# Velocity of excitations in ordered, disordered, and critical antiferromagnets

Arnab Sen,[1] Hidemaro Suwa,[2,3] and Anders W. Sandvik[3,4]

[1]*Department of Theoretical Physics, Indian Association for the Cultivation of Science, Jadavpur, Kolkata 700032, India*
[2]*Department of Physics, University of Tokyo, Tokyo 113-0033, Japan*
[3]*Department of Physics, Boston University, 590 Commonwealth Avenue, Boston, Massachusetts 02215, USA*
[4]*Department of Physics, National Taiwan University, Taipei 10607, Taiwan*

We test three different approaches, based on quantum Monte Carlo simulations, for computing the velocity $c$ of triplet excitations in antiferromagnets. We consider the standard $S = 1/2$ one- and two-dimensional Heisenberg models, as well as a bilayer Heisenberg model at its critical point. Computing correlation functions in imaginary time and using their long-time behavior, we extract the lowest excitation energy versus momentum using improved fitting procedures and a generalized moment method. The velocity is then obtained from the dispersion relation. We also exploit winding numbers to define a cubic space-time geometry, where the velocity is obtained as the ratio of the spatial and temporal lengths of the system when all winding number fluctuations are equal. The two methods give consistent results for both ordered and critical systems, but the winding number estimator is more precise. For the Heisenberg chain, we accurately reproduce the exactly known velocity. For the two-dimensional Heisenberg model, our results are consistent with other recent calculations, but with an improved statistical precision, $c = 1.65847(4)$. We also use the hydrodynamic relation $c^2 = \rho_s / \chi_\perp (q \to 0)$ between $c$, the spin stiffness $\rho_s$, and the transversal susceptibility $\chi_\perp$, using the smallest non-zero momentum $q = 2\pi/L$. This method also is well controlled in two dimensions, but the cubic criterion for winding numbers delivers better numerical precision. In one dimension, the hydrodynamic relation is affected by logarithmic corrections which make accurate extrapolations difficult. As an application of the winding number method, for the quantum-critical bilayer model our high-precision determination of the velocity enables us to quantitatively test, at an unprecedented level, the field-theoretic low-temperature scaling forms $\chi = (a/c^2)T$ and $C = (b/c^2)T^2$. We find agreement to within 3% with the leading $1/N$ results for the factors $a$ and $b$ in the O($N$) model, but the agreement becomes significantly worse when the first corrections in $1/N$ are considered.

## I. INTRODUCTION

Ordered quantum antiferromagnets exhibit linear dispersions of their elementary spin-wave (magnon) excitations and the associated velocity $c$ is an important parameter characterizing such systems. The prototypical two-dimensional Heisenberg model has an ordered ground state and its low-energy magnon spectrum is well described by spin-wave theory [1]. When $1/S$ corrections are properly taken into account, the velocity and other properties computed within this approximation for the most extreme (and interesting) case of the spin $S = 1/2$ model agree to within $\approx 1\%$ with the results of quantum Monte Carlo (QMC) calculations, which can be considered exact to within statistical errors if sufficiently large lattices are used for extrapolations to the thermodynamic limit [2–6]. This good agreement can be traced to the fact that the ground state is strongly ordered, the sublattice magnetization being reduced by quantum fluctuations by only about 40% from the classical value. Upon introducing other interactions, which enhance the quantum fluctuations and suppress the order, the quantitative predictive power of spin-wave calculations rapidly deteriorates and the quantum fluctuations have to be treated in more sophisticated ways [7–11]. An extreme case is when a system is driven to criticality. The low-energy critical excitations are still linearly dispersing but the corresponding velocity cannot be reliably calculated in any simple theoretical manner. Lastly, quantum-disordered antiferromagnets also have propagating triplet excitations, which are gapped and often called triplons.

In this paper, we will discuss three very different ways to extract the velocity of the elementary excitations of quantum spin models based on ground-state projector and finite-temperature QMC simulations. We consider one-dimensional (1D) and two-dimensional (2D) ordered, disordered, and critical quantum antiferromagnets. Using imaginary time dependent spin correlation functions, the long-time behavior computed at different momenta contain information on the dispersion relation, from which the velocity can be extracted if the limits of the system size going to infinity and the momentum going to zero are treated correctly. One can also in some cases extract the velocity in a simpler, indirect way using winding number fluctuations in the space-time representation of the system sampled in the QMC calculations [12,13]. We will develop a stable procedure based on these two approaches and compare the results for a number of different cases. In addition, we also test the well known hydrodynamic relationship $c^2 = \rho_s / \chi_\perp$ for an antiferromagnetic state [14], where $\chi_\perp$ is the transversal magnetic susceptibility and $\rho_s$ the spin stiffness.

We first successfully test the methods on both 1D and 2D systems for which the velocity (of spinons and spin waves, respectively) is previously known, and thereafter study a bilayer system, both in its paramagnetic phase and at its quantum-critical point. In the latter case, we subsequently use our high-precision estimate of $c$ to investigate in detail, to our knowledge at an unprecedented level of control, the reliability of finite-temperature quantum-critical scaling forms for the magnetic susceptibility and the specific heat [15,16].

Before turning to the calculations and results, in Sec. I A, we first provide some more background on the utility of carrying out precise determinations of the velocity in quantum spin systems, focusing on quantum-criticality in dimerized antiferromagnets. In Sec. I B, we provide a brief summary of the different calculations to be presented and in Sec. I C, we outline the organization of the rest of the paper.

### A. Effective low-energy descriptions

Quantum field theories are often used to describe universal low-energy properties of quantum magnets [15,16]. Numerical techniques, such as QMC simulations, can be used to extract system-dependent parameters appearing in various predicted forms of physical quantities at low energy. Such an approach of combining quantum field theory and numerics has been established over the past several years for certain types of quantum phase transitions in 2D systems [17,18]. Most well studied are transitions in *dimerized* models, where for weak interdimer coupling there is a tendency to singlet formation on the dimers, leading to a quantum phase transition into a quantum paramagnet at a critical ratio of the inter- and intradimer couplings [7,10,19–24]. A similar transition (of the same universality class) also take place if the dimers are replaced by some other unit cell of an even number of spins on which the single-cell ground state is a singlet [25,26]. The many studies of these systems have shown rather convincingly that the phase transition is in the 3D O(3) universality class, in agreement with field-theory descriptions based on the nonlinear sigma model [15,16]. Other properties associated with quantum criticality are also well captured by the field theory [16], e.g., the uniform magnetic susceptibility $\chi$ is linear in temperature at the critical coupling ratio, $\chi = aT$, and the specific heat grows quadratically; $C = bT^2$ at low $T$. Perhaps surprisingly, however, the values of the prefactors $a$ and $b$ have still not been tested quantitatively in a completely unbiased manner [27]. This may be largely because, as indicated above, they depend on the velocity $c$ of the critically damped spin waves (and on no other low-energy parameter), but this parameter has not yet been independently calculated to high precision for model systems (by first-principles methods not depending on other field-theory predictions). As a demonstration of the value of determining $c$ to high precision, in this paper, we will also provide a test case of a detailed comparison with field-theory predictions for one of the prototypical dimerized models; the Heisenberg bilayer.

### B. Technical and physics objectives

The first aim of the present paper is to systematically test three completely different ways of extracting the velocity of elementary excitations based on QMC calculations in the following ways: (i) using the momentum-dependent gaps extracted from imaginary time-dependent spin correlation functions and (ii) by monitoring spatial and temporal winding-number fluctuations, which are proportional to the spin stiffness and the uniform susceptibility, respectively, and adjusting the space-time aspect ratio $L/\beta$ such that these fluctuations are equal. At this special inverse temperature $\beta^*(L)$, the ratio of the spatial and temporal lengths $L/\beta^*$ should equal $c$ [12,13].

(iii) The third way is by exploiting the hydrodynamic relation $c^2 = \rho_s/\chi_\perp$, where one has pay attention to the fact that $\chi_\perp \to 0$ when the temperature $T \to 0$ in a finite system, while the spin stiffness remains nonzero. We discuss an approach to circumvent this problem based on $\chi_\perp(q)$ where the momentum $q$ is small but nonzero.

The method (i) is in principle very direct, being connected to the fundamental definition of $c$ in the dispersion relation of the lowest-energy excitation versus momentum. However, the precise determination of the momentum-dependent gap is in practice complicated by the presence of a continuum of excitations above the lowest energy. In some cases, as we will show, one also has to take great care with the way the thermodynamic limit and zero-momentum limits are taken. The method (ii) is rather simple and the only uncertainty is introduced by a final extrapolation of the finite-size velocity definition $L/\beta^*$ to the thermodynamic limit. However, as far as we are aware, the correctness of this approach has not been formally proven, except for the case of a long-range ordered antiferromagnet, and the method has not been widely used [12,13]. We here confirm that the method continues to give the correct velocity also when the system is critical, even in the case of the $S = 1/2$ Heisenberg chain, where the low-energy excitations are not even spin-waves but deconfined spin-1/2 carrying topological defects ("spinons"). The method (iii) based on generalized hydrodynamics is simple to apply and we will argue that it as well continues to work also in critical systems.

Comparing methods (i)–(iii) for the Heisenberg chain as well as for the well-studied 2D Heisenberg model gives us insights into how to best apply the methods in practice. After these tests, we study the bilayer Heisenberg model at its quantum-critical point and in its paramagnetic regime, using methods (i) and (ii). Also in this case we find good agreement between the two methods at the critical point, but one has to be more careful when defining the velocity based on gaps for finite systems, because of a slower convergence of the gaps to their infinite-size values than in the ordered state. We also explicitly show that the winding number estimator gives an *incorrect* velocity in the gapped paramagnetic phase.

The second aim of the paper connects to the low-energy filed-theory description discussed in Sec. I A—to compute a high-precision value for $c$ for the quantum-critical Heisenberg bilayer model and to use this value to reliably test the field-theory predictions for $\chi(T)$ and $C(T)$. While many such tests have been performed in the past, on the bilayer [10,19,21] as well as other [25] quantum-critical 2D spin systems, the past studies were not able to completely quantitatively test the level of agreement with the existing large-$N$ field-theory predictions, because an independent, unbiased determination of $c$ was lacking. This obstacle is overcome by the reliable calculation of $c$ in this paper.

### C. Outline of the paper

The rest of the paper is organized as follows. In Sec. II, we discuss calculations of imaginary-time correlation functions within a ground-state projector QMC method formulated in the basis of valence bonds. We discuss our methods to extract the lowest momentum-dependent gaps from the long-time

behavior of these correlations. In addition to fitting a leading exponential decay with additional corrections to account for higher excitations, we also present a generalization to ground-state projection QMC data of a systematic high-order moment approach recently introduced for use with finite-temperature QMC simulations [28]. We compare our results with exact solutions for small Heisenberg chains as well as with the rigorously known velocity of this model in the thermodynamic limit. In Sec. III, we discuss the determination of $c$ based on winding numbers and demonstrate the method using the Heisenberg chain. These finite-temperature calculations are carried out with the stochastic series expansion (SSE) QMC method. In Sec. IV, we discuss details of the hydrodynamic relationship generalized to finite system size and test it using SSE calculations for the Heisenberg chain. In Sec. V, we present further tests of all three methods for the standard 2D Heisenberg model. We also show that the scaling of the triplet gap at momentum $\mathbf{k} = (\pi, \pi)$ is entirely consistent with quantum rotor-states carrying spin $S = 1$. In Sec. VI, we discuss results for the critical and disordered bilayer systems, including the comparison of $c$ determined using both methods. The analysis of the quantum-critical susceptibility and specific heat computed using large-scale SSE calculations is presented in Sec. VII. We briefly summarize and discuss our results further in Sec. VIII.

## II. MOMENTUM-DEPENDENT SPIN GAPS

Here, we discuss how to use zero temperature ($T = 0$) projector QMC methods to calculate imaginary-time correlation functions, which are then used to extract the triplet gap (with respect to the ground-state energy) as a function of momentum $\mathbf{k}$. Applying the imaginary-time evolution operator $e^{-\beta H}$ to a trial state leads to the ground state when $\beta \to \infty$. Alternatively, one can use a high power $H^m$ of the Hamiltonian and reach the ground state when $m \to \infty$. In practice, QMC simulation methods based on these two operators are very similar, but the exponential form allows more direct access to the standard imaginary time. As we will show, this connection is not needed if only the excitation energies are of interest, and then one can use the slightly faster power method. We will use both approaches here, with two different ways of analyzing the correlation functions.

### A. $H$-power projection

This QMC approach is based on projection with a sufficiently high power of the Hamiltonian $H$ on a trial state $|\psi_t\rangle$. The process can be conveniently expressed in the energy eigenbasis of $H$, leading to the following expression for the dependence on $m$:

$$(-H)^m |\psi_t\rangle = c_0 (-E_0)^m \left[ |0\rangle + \sum_{n=1}^{\Lambda} \frac{c_n}{c_0} \left( \frac{E_n}{E_0} \right)^m |n\rangle \right]. \quad (1)$$

Here, $|n\rangle$, $n = 0, \ldots, \Lambda - 1$ are the energy eigenstates of $H$ and $E_0$ is assumed to be negative, with its absolute value $|E_0|$ being the largest in magnitude of all the energies (which can always be achieved by adding a suitable constant to the Hamiltonian). Then, if the expansion coefficient $c_0 \neq 0$ (which in practice is essentially guaranteed for any reasonable

choice of $\psi_t\rangle$), we have $(-H)^m |\psi_t\rangle \propto |0\rangle$ for large $m$, and the expectation value of an operator $O$ at $T = 0$ can be written as

$$\langle O \rangle = \frac{\langle \psi_t | (-H)^m O (-H)^m | \psi_t \rangle}{\langle \psi_t | (-H)^{2m} | \psi_t \rangle}. \quad (2)$$

For the SU(2) invariant spin models considered in this paper, this form of the expectation value can be evaluated by importance-sampling, using a formulation of the projector QMC method in the nonorthogonal valence bond basis [29]. An efficient way of sampling the contributions to $(-H)^{2m}$, very similar to "operator-loop" updates developed within the finite-$T$ SSE method [30], can be formulated using loop updates in a combined space of spin components ($S^z$) and valence bonds [5]. The trial state is also expressed using valence bonds, in the form of an amplitude-product state [31]. The details of the state (the form of the amplitudes) are not important, as the good convergence to the ground state is observed even if the state is not optimized [5].

We will use this valence-bond variant of the projector QMC method for computing appropriate imaginary-time dependent correlation functions. For technical details on the sampling methods we refer to Ref. [5]. Below we will focus on the definition of the correlation functions we study and how we process them to extract the velocity.

#### 1. Imaginary-time correlations

We consider correlation functions of the following form:

$$C_A(t) = \frac{\langle 0 | A^\dagger (-H)^t A | 0 \rangle}{\langle 0 | (-H)^t | 0 \rangle} = \sum_{n=0}^{\Lambda-1} \left| \frac{E_n}{E_0} \right|^t |\langle n | A | 0 \rangle|^2, \quad (3)$$

where $t$ is an integer, which can be related to imaginary time [32] and we will loosely refer to it by this term. More precisely, $t/N$, where $N$ is the system size, is proportional to imaginary time $\tau$ in the sense of the standard Schrödinger time evolution operator $e^{-\tau H}$, as we will explicitly show below. From $C_A(t)$, we further define

$$Q_A(t) = \frac{C_A(t) - |\langle 0 | A | 0 \rangle|^2}{C_A(0) - |\langle 0 | A | 0 \rangle|^2}, \quad (4)$$

and note that $Q_A(0) = 1$ and $Q_A(t \to \infty) \to 0$. For large $t$, we have

$$Q_A(t) \to \left( \frac{|\langle 1 | A | 0 \rangle|^2}{\langle 0 | A^\dagger A | 0 \rangle - |\langle 0 | A | 0 \rangle|^2} \right) \left( 1 - \frac{\Delta}{|E_0|} \right)^t, \quad (5)$$

where $\Delta = E_1 - E_0$ is the energy gap between the first excited state (connected to the ground state by the operator $A$) and the ground state. To directly show the relationship between $t$ and imaginary time, we can introduce $\tau = t/|E_0|$, then write $E_0 = N e_0$, and in the limit of large $N$ have

$$Q_A(\tau = t/|E_0|) \propto \left( 1 - \frac{\Delta}{N|e_0|} \right)^{N|e_0|\tau} \to e^{-\Delta \tau}, \quad (6)$$

which is the familiar form of the asymptotic decay of an imaginary-time dependent correlation function. Thus, we have shown that, indeed, $\tau \propto t/N$. We will not explicitly need to make use of the relationship between $t$ and $\tau$ here, however, and we will continue to use $t$ as the "time" parameter with the $H$-power method.

We can appropriately choose the operator $A$ so that it excites the ground state $|0\rangle$ into a state with desired quantum numbers. The ground state of the unfrustrated Hamiltonians considered here are total spin singlets with momentum $\mathbf{k} = 0$ on a finite lattice with an even number of spins and periodic boundary conditions (in one dimension only when the number $N$ of sites is a multiple of four—for other even $N$ the momentum is $\pi$) [33]. Since we are interested in triplet excitations, we can use the simple Fourier-transformed spin operator,

$$A(\mathbf{k}) = S^z(\mathbf{k}) = \frac{1}{\sqrt{N}} \sum_{\mathbf{r}} e^{i\mathbf{k}\cdot\mathbf{r}} S^z(\mathbf{r}), \qquad (7)$$

to create an $S = 1$ state with momentum $\mathbf{k}$ and $S^z = 0$ when acting on the ground state. Thus the following imaginary-time correlation function

$$\begin{aligned} C_{\mathbf{k}}(t) &= \frac{\langle 0|A(-\mathbf{k})(-H)^t A(\mathbf{k})|0\rangle}{\langle 0|(-H)^t|0\rangle} \\ &= \frac{\langle \psi_t|(-H)^{2m-p-t} A(-\mathbf{k})(-H)^t A(-H)^p|\psi_t\rangle}{\langle \psi_t|(-H)^{2m}|\psi_t\rangle} \end{aligned} \qquad (8)$$

allows us to directly measure the triplet excitation gap $\Delta(\mathbf{k})$ as a function of the momentum. The second line in the above equation explicitly shows the form used in the projector QMC calculations, where both $2m - p - t$ and $p$ are assumed to be large enough to achieve projection to the ground state for the system sizes considered.

In practice, we will use projection powers $m = 16 p_r N$ and, to achieve good ground-state convergence, require that $2m - p - t$ and $p$ both are larger than $15 p_r N$, where typically $p_r = 8$ (16 or higher in some cases). These choices are motivated by extensive tests indicating that no detectable systematical errors remain. The values of $t$ are restricted to be multiples of $N/4$ in our simulations. Since the correlation functions at the different $t$ values are measured in the same simulation, the $C_{\mathbf{k}}(t)$ data are correlated and measuring at shorter $t$ intervals does not significantly increase the amount of statistical information in the data set.

### 2. Extracting the gap

We here use two different ways to extract the lowest triplet gap from the correlation function $C_{\mathbf{k}}(t)$. Since $\langle 0|A(\mathbf{k})|0\rangle = 0$ for $\mathbf{k} \neq 0$, Eq. (4) reduces to

$$Q_{\mathbf{k}}(t) = \frac{C_{\mathbf{k}}(t)}{C_{\mathbf{k}}(0)}. \qquad (9)$$

In principle, the gap can be extracted by monitoring the long-time behavior, given by the form (5). A systematic way to extract the gap without performing any curve fits is to consider the ratio of $Q_{\mathbf{k}}(t)$ at two different times separated by some interval; e.g., by $N/4$ operations:

$$R_{\mathbf{k}}(t) = |E_0| \left\{ 1 - \left[ \frac{Q_{\mathbf{k}}(t + N/4)}{Q_{\mathbf{k}}(t)} \right]^{4/N} \right\}. \qquad (10)$$

Note that $R_{\mathbf{k}}(t) \to \Delta(\mathbf{k})$ when $t$ is large enough, which follows from the long-time behavior of $Q_{\mathbf{k}}(t)$ in Eq. (5). However, in our QMC calculations, it is not always possible to reach perfect convergence of this gap estimate for all $\mathbf{k}$, because the relative statistical errors often become too large already for moderately

large $t$. This problem is related to the existence of a continuum (for large $N$) of states above the gap, due to which the pure exponential decay cannot be easily observed in practice.

We thus use another method to estimate the value of the gap based on the entire available set of correlation functions. This scheme is more reliable than the ratio scheme when data (with small relative errors) are not available for large values of imaginary time. It is clear from Eq. (3) that one can define a positive-definite spectral function $A_{\mathbf{k}}(\omega)$ to fit the normalized imaginary-time correlation $Q_{\mathbf{k}}(t)$ as

$$Q_{\mathbf{k}}(t) = \int_0^\infty d\omega A_{\mathbf{k}}(\omega) \left( 1 - \frac{\omega}{|E_0|} \right)^t. \qquad (11)$$

This is just the analog for the $H$-power evolution of the standard form,

$$G_{\mathbf{k}}(\tau) = \int_0^\infty d\omega S_{\mathbf{k}}(\omega) e^{-\omega\tau}, \qquad (12)$$

relating the imaginary-time Schrödinger evolved correlation function

$$G_{\mathbf{k}}(\tau) = \langle 0|S^z_{-\mathbf{k}}(\tau) S^z_{\mathbf{k}}(0)|0\rangle, \qquad (13)$$

to the dynamic spin structure factor $S_{\mathbf{k}}(\omega)$. In Eq. (11), $\omega$ cannot exceed $|E_0|$, following from the fact that we have ensured that $E_0$ is the eigenvalue with the largest magnitude. In practice, as in the standard dynamic spin structure factor in (12), the actual dominant spectral weight will be concentrated only to within a window of order $J$.

For any finite system, $A_{\mathbf{k}}$ is a sum of $\delta$ functions and this can be replaced by a continuum starting at $\omega = \Delta(\mathbf{k})$ for a large system (or, in some cases, there is an isolated $\delta$ function at the gap, following by a second gap and then a continuum). With $G_{\mathbf{k}}(\tau)$ or $Q_{\mathbf{k}}(t)$ computed using QMC calculations, the respective relations (11) or (12) can in principle be inverted using numerical analytic continuation. This procedure is very challenging, however, and it is not easy to extract the gap precisely with conventional methods such as the maximum entropy method [34], though one can extract the main dominant spectral features (and we note that progress in this regard has been made very recently [35]). Here our goal is merely to extract the gap, and instead of trying to reproduce the full spectral function we model the excitations by just a small number of $\delta$ functions. With the precision of typical QMC data, $Q_{\mathbf{k}}(t)$ can be normally fitted very well with just a few $\delta$ functions (typically 3 to 5) over the full range of accessible times $t$. With this procedure, we expect the location $\omega_1$ of the lowest gap to accurately reproduce $\Delta(\mathbf{k})$, while the higher $\delta$ functions represent approximately the contributions of the continuum. In this fitting procedure, the extracted $\omega_1$ is to some extent affected by contributions of the higher states but does not change significantly when increasing the number of $\delta$ functions. We can therefore quite reliably extract the lowest gap, but not higher ones unless they are separated by significant subsequent gaps (which is not expected in the cases of interest here, except well inside the quantum paramagnetic state of the bilayer model).

Given a set of $n$ $\delta$ functions at energies $\omega_i$ with associated amplitudes $A_i$ normalized so that $\sum_i A_i = 1$, one can compute the associated time dependent correlation function in analogy

with Eq. (11) as

$$Q_{\mathbf{k}}(t) = \sum_{i=1}^{n} A_i \left(1 - \frac{\omega_i}{|E_0|}\right)^t. \tag{14}$$

Now denoting the corresponding QMC-computed function by $\tilde{Q}_{\mathbf{k}}(t)$ and their statistical error by $\sigma_t$, the goodness of the fit is quantified in the standard way by $\chi^2$, based on a set of $N_t$ time points $\{t\}$:

$$\chi^2 = \frac{1}{N_t} \sum_{\{t\}} \frac{1}{\sigma_t^2} (Q_{\mathbf{k}}(t) - \tilde{Q}_{\mathbf{k}}(t))^2. \tag{15}$$

We here use a uniform grid of time points with separation $N/4$ operations, $t = N/4, N/2, \ldots$, up to a point $t_{\max} = N_t(N/4)$ where the relative statistical error of $\tilde{Q}_{\mathbf{k}}(t)$ exceeds 5%. The choice of cutoff is not very important as, in any case, the noisy data at very large $t$ will not affect the fit from the definition of $\chi^2$.

For the extracted gap to be reliable, the contribution of the lowest $\delta$ function to the fit must be significant at the longest times included. To monitor this long-time weight, we compute the relative contribution of the lowest $\delta$ function, denoted by $A_1(t)$, at the time $t_{\max}$ included in the fit:

$$A_1(t_{\max}) = \frac{A_1(1 - \omega_1/|E_0|)^{t_{\max}}}{\sum_{i=1}^{n} A_i (1 - \omega_i/|E_0|)^{t_{\max}}}. \tag{16}$$

This quantity should approach 1 for $t \to \infty$ if the lowest $\delta$ function is at the gap. It is close to 1 in all the fits reported here, indicating stable extraction of $\omega_1$.

The statistical error of the extracted gap is estimated using a bootstrap error analysis. With the underlying QMC data for the correlation function $C_{\mathbf{k}}(t)$ saved as $M$ bin averages (with typically $M \sim 100$–$1000$), bootstrap averages are constructed by selecting $M$ bins at random (i.e., allowing for the same bin to be selected multiple times). The above fitting procedures are then carried out repeatedly for a large number of these samples, and the standard deviation of the estimates is the statistical error of the gap in our procedure.

As already mentioned, the fluctuations of the QMC data at different times are significantly correlated since these are measured in the same simulation. A statistically correct treatment of the data would require the inclusion of the full covariance matrix (instead of just its diagonal elements) in the definition of $\chi^2$. However, much of the covariance is already removed when the time-correlations are normalized [by the denominator Eq. (9)], because the errors are correlated primarily by overall fluctuations in the normalization. Based on test cases, including ones reported below, to obtain fully reliable results it is sufficient to use only the diagonal elements of the covariance matrix and define $\chi^2$ as in (15).

### 3. Tests on the Heisenberg chain

We here illustrate the gap extraction method described in the previous subsection using the example of the $S = 1/2$ Heisenberg spin chain with periodic boundary conditions, where spins interact with nearest neighbor exchange

constant $J = 1$;

$$H = J \sum_{i=1}^{L} \mathbf{S}_i \cdot \mathbf{S}_j. \tag{17}$$

First, we compare the results of our numerics for the lowest triplet gap at $k = \pi$ for chain sizes $L = 12$ and $24$ with exact diagonalization (Lanczos-method) results. As can be seen in Fig. 1(a), the quantity $R_\pi(t)$ defined in Eq. (10) indeed converges to the correct gap value in both the cases. Figure 1(b) shows the normalized imaginary time correlation function $Q_\pi(t)$ for the two system sizes, and Fig. 1(c) shows the distribution of the gap error obtained using a bootstrap method (i.e., the simulation data stored as $M$ bin averages are resampled by selecting $M$ bins at random, and the fitting procedures are carried out for each such averaged data set) for $L = 24$, using a fit to three $\delta$ functions.

This analysis show that the gap obtained from the fit agrees statistically with the exact result, which here is within a standard deviation of the distribution obtained in the bootstrapping procedure, and the distribution itself closely matches a normal distribution. Thus, even with only three $\delta$ functions in the spectrum (which is clearly a much smaller number than what is contained in the full spectrum), we detect no systematic errors introduced by the fitted functional form, supporting our assertion that the simplified description of the spectrum does not significantly affect the location of the lowest $\delta$ function (the gap). The number $n$ of $\delta$ functions that should be included in a given case depends on the statistical errors of the QMC data and the actual form of the spectral continuum. To determine $n$, we monitor $\chi^2$ as a function of increasing $n$ and stop when no improvements in the fit are observed.

For the Heisenberg chain, the lower edge of the spectrum of triplet excitations is known rigorously based on the Bethe ansatz solution [36]. For momentum $k \to 0$ and $k \to \pi$, the spectrum is linear with velocity $c = \pi/2$. For an infinite chain, the triplets are degenerate with singlet excitations, due to the fact that the excitations consist of essentially independently propagating deconfined spinons, each carrying spin $1/2$. The 2-spinon continuum is maximally broad at $k = \pi$ and shrinks to zero at $k = 0$. This leads to larger contributions to the time correlations from the continuum close to $k = \pi$, which is directly visible in the QMC data for long chains as, e.g., a slower rate of convergence of $R_k(t)$ to a constant. For example, $R_{2\pi/L}(t)$ converges much faster to its asymptotic constant value compared to $R_\pi(t)$ for a chain of size $L = 64$, as shown in Fig. 2. In the inset of the same figure, we also show results for $k = \pi + 2\pi/L$, which is the momentum we use to extract $c$ as discussed further below. Here the chain is longer, $L = 200$, and $R_k(t)$ does not converge sufficiently before the error bars become too large. However, the method of fitting a simplified spectral function to $Q_\pi(t)$ still works very well and delivers a gap consistent with $R_k(t)$, but with much better statistical precision. We therefore exclusively uses the fitting method to obtain the results to be discussed next.

In a finite chain, the ground state is nondegenerate and has momentum $k = 0$. However, there is also a quasidegenerate state with $k = \pi$, which is obtained by adding an umklapp to the true ground state, and this state becomes exactly degenerate with the $k = 0$ ground state when $L \to \infty$ [37]. There is also

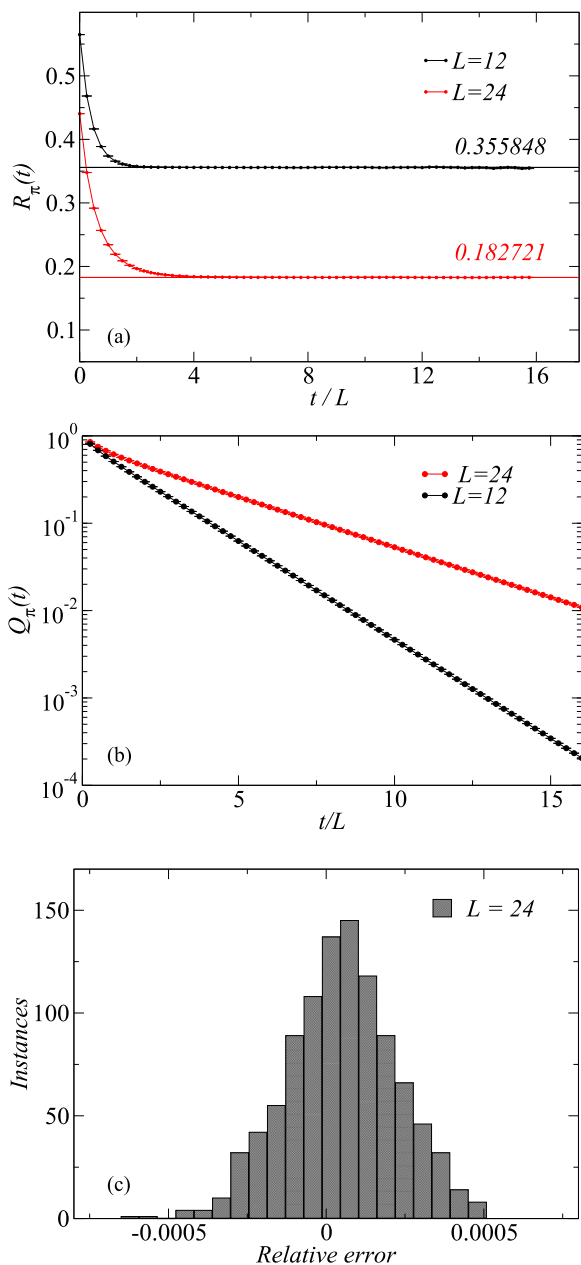FIG. 1. (Color online) (a) Correlation ratio $R_\pi(t)$, Eq. (10), shown for two Heisenberg chain sizes, $L = 12$ and 24. The lines are the exact values of the triplet gap at this wave-vector (with the numerical values indicated as well). (b) The normalized correlation function $Q_\pi(t)$ for $L = 12$ and 24. (c) The distribution of the relative error of the gap $\Delta(\pi)$ extracted by fitting the data for $L = 24$ in (b) using three $\delta$ functions. The histogram was generated from a large number of bootstrap samples of the QMC data. The relative error is calculated as $(\Delta - \Delta_e)/\Delta_e$, where $\Delta_e$ is the exact value of the gap. The exact value (corresponding to relative error 0) is seen to fall within a standard deviation of the distribution of the bootstrap samples.

a corresponding finite-size shift in the excitations close to $k = \pi$, which is characteristic of the Heisenberg chain but not present in the model in higher dimensions. The lowest triplet gap in the neighborhood of $k = \pi$, where we will use only
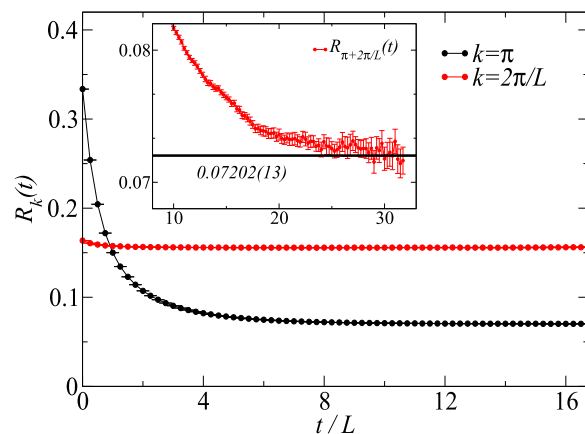


FIG. 2. (Color online) Convergence of the correlation ratio $R_k(t)$ to the gap $\Delta(\mathbf{k})$ as a function of imaginary time $t$, shown at $k = \pi$ and $2\pi/L$ for a Heisenberg chain of length $L = 64$. The inset shows $R_k(t)$ at $k = \pi + 2\pi/L$ for a longer chain, $L = 200$, where the gap extraction using a fit of $Q_k(t)$ to five $\delta$ functions works demonstrably better, giving the result indicated below the horizontal line, with the thickness of the line representing the average gap $\pm$ one error bar.

$k = \pi + 2\pi/L$, behaves for large $L$ as [38]

$$\Delta(\pi + 2\pi/L) = \Delta(\pi) + c(L)\frac{2\pi}{L}, \tag{18}$$

where $\Delta(\pi) \sim 1/L$ but with a multiplicative logarithmic correction. In Fig. 3, we show the behavior of the corresponding velocity estimate,

$$c(L) = \frac{L}{2\pi}[\Delta(\pi + 2\pi/L) - \Delta(\pi)], \tag{19}$$

as a function of the inverse chain length. A smooth monotonic (asymptotically linear) approach to the known velocity, $c = \pi/2$ can be observed as $L \to \infty$. There are no signs of any remaining logarithmic corrections as a function of $1/L$ in this estimate. Our results are consistent with a remaining linear finite-size correction.
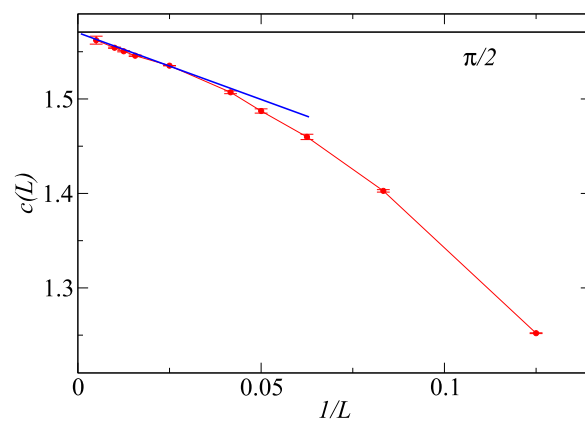


FIG. 3. (Color online) Finite-size velocity estimates for the Heisenberg spin chain obtained from triplet gaps at and in the neighborhood of $k = \pi$ according to Eq. (19), shown as a function of the inverse system size. The velocity approaches the known value $c = \pi/2$ (indicated by the horizontal line) for large systems. The line through the larger system sizes is a guide to the eye.
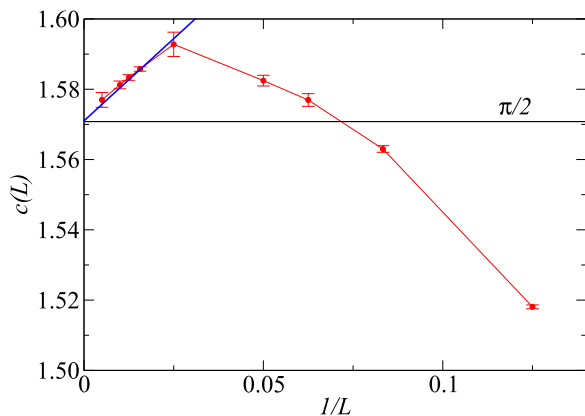
FIG. 4. (Color online) Velocity estimate for the Heisenberg chain from the triplet gap close to $k = 0$; $c(L) = \Delta(k)/k$, where $k = 2\pi/L$, shown as a function of the inverse system size. The known velocity $c = \pi/2$ in the thermodynamic limit is indicated by the horizontal line. The line through the larger system sizes is not a fit but is drawn to match the known velocity and the data for those system sizes.

The velocity of the spinons can also be extracted from the lowest triplet gap at $k = 2\pi/L$ by using the simple estimator $L\Delta(2\pi/L)/(2\pi) \rightarrow c$ as $L \rightarrow \infty$. Results are shown in Fig. 4. Also in this case we observe the estimate approaching the correct value of $c$ as $L \rightarrow \infty$, but with a nonmonotonic behavior with a maximum for $L \approx 40$ before an apparently linear asymptotic approach to the correct value. It should be noted that the spectral weight (before normalizing the correlation function) vanishes as $k \rightarrow 0$, which implies that the $k = 2\pi/L$ correlation function computed in the QMC simulations becomes very noisy for large system sizes. It is worth noting here that, because of the nonmonotonic behavior seen in Fig. 4, a calculation using exact (numerical) diagonalization of the Hamiltonian would appear inconsistent with the known velocity, because large enough system sizes required to go beyond the maximum at $L \approx 40$ cannot be reached.

### B. Exponential-form projection

In this section, we will discuss a different way of extracting the gap, using a systematic way to analyze moments of the spectrum based on information contained in the imaginary-time correlations. Such a scheme was recently introduced within $T > 0$ QMC calculations [28], and we here present a generalization to projector QMC. In addition, we discuss improvements in the extrapolations required to obtain unbiased results.

To make the connection with the previous version of the method more transparent, we will here use the exponential-form projection with continuous imaginary time. The ground state is again projected out of a trial state in the valence-bond basis, but now using $e^{-\beta H}$ instead of $(-H)^m$. Taylor expanding the exponential (as was also done previously in projector QMC, e.g., in Ref. [39]), the scheme then closely resembles $T > 0$ SSE QMC algorithms [30]. The normalization, which replaces the partition function in $T > 0$ methods, is expressed

as

$$
\begin{aligned}
Z' &= \langle \psi_t | e^{-\beta H} | \psi_t \rangle \\
&= \langle \psi_t | T_\tau \exp\left( -\int_0^\beta d\tau\, H(\tau) \right) | \psi_t \rangle \\
&= \langle \psi_t | 1 + \sum_{m=1}^\infty (-1)^m \int_0^\beta d\tau_1 \int_{\tau_1}^\beta d\tau_2 \cdots \\
&\quad \cdots \int_{\tau_{m-1}}^\beta d\tau_m \prod_{i=1}^m H(\tau_i) | \psi_t \rangle,
\end{aligned}
\tag{20}
$$

where $T_\tau$ indicates time ordering and $H(\tau)$ is a Hamiltonian acting at imaginary time $\tau$. If the Hamiltonian $H(\tau)$ here is actually time dependent, we obtain the nonequilibrium QMC scheme developed in Ref. [40]. Here $H$ is time independent, and we formally use the time-dependent formalism only as a convenient way of accessing imaginary time in the configuration space directly. In principle, we can also use the time dependence corresponding to the interaction representation [41–43], where only the $x$ and $y$ parts of the interaction appears in the operator product, but here we stay close to the formulation also used in the SSE method and in the $H$-power method discussed above, and expand with the full Hamiltonian. Note that the $H$-power method is exactly recovered if the time integrals are completed, and one can in practice also equivalently use the power method and just generate the ordered time sequences at random [43], or generate them in the process of the updates as in the world-line continuous time algorithm [41].

#### 1. Generalized moment method

Our approach introduced here to extract the gap is a generalization to projection QMC of the moment method proposed recently [28]. The Fourier-transformed dynamical correlation is exploited to derive a sequence of gap estimators. We will first explain how the moment of the dynamical correlation function is related to the gap, and then derive increasingly precise gap estimators. Finally, we will show how these estimates are extrapolated to the limit in which the exact gap is recovered.

We will use the following dynamical correlation function computed with the exponential projector:

$$
\begin{aligned}
C(\tau, \tau_0) &= \frac{1}{Z'} \langle \psi_t | e^{-(\beta - \tau - \tau_0)H} A^\dagger e^{-\tau H} A e^{-\tau_0 H} | \psi_t \rangle \\
&= \frac{1}{Z'} \sum_{\ell, p, q} b_{\ell, p, q}\, e^{-\beta E_p} e^{-\tau(E_\ell - E_p)} e^{-\tau_0(E_q - E_p)} \\
&\rightarrow \sum_{\ell \geqslant 1} b_\ell e^{-\tau \Delta_\ell} \quad (\beta, \tau_0 \rightarrow \infty),
\end{aligned}
\tag{21}
$$

where we use the definitions

$$
b_{\ell, p, q} = \bar{c}_p c_q \langle p | A^\dagger | \ell \rangle \langle \ell | A | q \rangle, \tag{22}
$$

$$
b_\ell \equiv b_{\ell, 0, 0} = |c_0 \langle \ell | A | 0 \rangle|^2, \tag{23}
$$

$$
\Delta_\ell = E_\ell - E_0. \tag{24}
$$

We assume that $A|0\rangle \neq 0$, $\langle 0|A|0\rangle = 0$, and $\Delta_1 > 0$, all of which apply here.

Let us first consider the moment of the asymptotic correlation function in the limit $\beta, \tau_0 \to \infty$:

$$
\begin{aligned}
I_n^\infty &= \int_0^\infty d\tau \, \tau^n C(\tau, \tau_0) \\
&= \int_0^\infty d\tau \, \tau^n \sum_{\ell \geqslant 1} b_\ell \, e^{-\tau \Delta_\ell} \\
&= \sum_{\ell \geqslant 1} \frac{b_\ell}{\Delta_\ell^{n+1}} n! \\
&\sim \frac{b_1}{\Delta_1^{n+1}} n! \quad (n \gg 1).
\end{aligned}
\tag{25}
$$

Then the lowest gap can, in principle, be obtained using an appropriate ratio, e.g.,

$$
\frac{(n+1)I_n^\infty}{I_{n+1}^\infty} \to \Delta_1 \quad (n \to \infty).
\tag{26}
$$

In practice, however, the projection time $\beta$ cannot be infinite in simulations, and we have to consider carefully the effects of finite $\beta$. In addition, the range of integration, $\tau \in (0, \beta_{\text{int}})$, is different from (less than) $\beta$ because the imaginary-time correlations are measured from the reference point $\tau_0 = \beta/2$ at the center of the projected trial state [noting that Eq. (20) corresponds to projection of the bra and ket state with $e^{-\beta/2}$], or, alternatively, between two points located symmetrically within the projection range $\tau \in (0, \beta_{\text{int}})$, and one has to stay well away from the boundaries (trial states) for unbiased measurements. Moreover, in many cases, the statistical errors grow too large at large times, as mentioned previously. Thus, in practice, $\beta_{\text{int}} < \beta/2$.

The moments for finite $\beta$ and $\beta_{\text{int}} \leqslant \beta/2$ take the form

$$
\begin{aligned}
I_n &= \int_0^{\beta_{\text{int}}} d\tau \, \tau^n C(\tau, \tau_0) \\
&= \frac{1}{Z'} \int_0^{\beta_{\text{int}}} d\tau \, \tau^n \sum_{\ell, p} d_{\ell, p} e^{-\tau \Delta_{\ell, p}} \\
&= \frac{1}{Z'} \sum_{\ell, p} \frac{d_{\ell, p} \, n!}{\Delta_{\ell, p}^{n+1}} \left( 1 - \sum_{m=0}^n P(m) \right),
\end{aligned}
\tag{27}
$$

where

$$
d_{\ell, p} = e^{-\beta_{\text{int}} E_p} \sum_q b_{\ell, p, q} e^{-\tau_0 (E_q - E_p)},
\tag{28}
$$

$$
\Delta_{\ell, p} = E_\ell - E_p,
\tag{29}
$$

and

$$
P(m) = \frac{(\beta_{\text{int}} \Delta_{\ell, p})^m e^{-\beta_{\text{int}} \Delta_{\ell, p}}}{m!}
\tag{30}
$$

is a properly normalized Poisson distribution;

$$
\sum_{m=0}^\infty P(m) = 1.
\tag{31}
$$

Owing to the finite integration range, the Poisson term does not vanish. As a consequence, the ratio of the moments for finite

$\beta$ does not contain information of the gap in the limit $n \to \infty$. Instead, we have a completely different limiting behavior,

$$
\frac{I_n}{I_{n+1}} \sim \frac{n+2}{\beta_{\text{int}}(n+1)} \to \frac{1}{\beta_{\text{int}}} \quad (n \to \infty),
\tag{32}
$$

independent of the gap. We can overcome this difficulty and devise a proper gap estimator by using the Fourier transformation. Here we express the moments in a different way using the following expansion:

$$
\frac{I_{2n}^\infty}{(2n)!} = \lim_{\beta_{\text{int}}, \tau_0 \to \infty} \frac{(-1)^n}{\omega_1^{2n}} \sum_{k=0}^n x_{n,k,0} R(\omega_k),
\tag{33}
$$

$$
\frac{I_{2n-1}^\infty}{(2n-1)!} = \lim_{\beta_{\text{int}}, \tau_0 \to \infty} \frac{(-1)^{n-1}}{\omega_1^{2(n-1)}} \sum_{k=1}^n x_{n,k,1} \frac{J(\omega_k)}{\omega_k},
\tag{34}
$$

where $R(\omega_k)$ and $J(\omega_k)$ are the real and imaginary parts, respectively, of the Fourier-transformed correlation function, i.e.,

$$
\int_0^{\beta_{\text{int}}} d\tau \, e^{i \tau \omega_k} C(\tau, \tau_0) = R(\omega_k) + i J(\omega_k),
\tag{35}
$$

where $\omega_k = 2\pi k / \beta_{\text{int}}$ ($k \in \mathbf{Z}$), and the key coefficients are $x_{1,1,1} = 1$ and

$$
x_{n,k,m} = \frac{1}{\prod_{m \leqslant j \leqslant n, j \neq k} (k+j)(k-j)}.
\tag{36}
$$

When deriving these equations we have considered the expansion in $\omega_k$ on the right-hand side of Eqs. (33) and (34), where the lowest orders then cancel. The coefficients $x_{n,k,m}$ can be solved for by using the inverse of the Vandermonde matrix [44].

Combining the results above, we obtain the following improved gap estimator:

$$
\hat{\Delta}_{(n, \beta_{\text{int}})} = -\omega_1^2 \frac{\sum_{k=1}^n x_{n,k,1} \frac{J(\omega_k)}{\omega_k}}{\sum_{k=0}^n x_{n,k,0} R(\omega_k)}.
\tag{37}
$$

Remarkably, this estimator is asymptotically unbiased and the limits are interchangeable (see Appendix for the analytical derivation):

$$
\begin{aligned}
\Delta_1 &= \lim_{n \to \infty} \lim_{\beta_{\text{int}}, \tau_0 \to \infty} \hat{\Delta}_{(n, \beta_{\text{int}})} \\
&= \lim_{\beta_{\text{int}}, \tau_0 \to \infty} \lim_{n \to \infty} \hat{\Delta}_{(n, \beta_{\text{int}})}.
\end{aligned}
\tag{38}
$$

Due to the commuting limits, observing convergence of the estimator (37) in large $n$ and $\beta_{\text{int}}$ taken in any convenient fashion will deliver an unbiased result for the gap.

In principle, there is also a dependence on the $\beta$ value used in the exponential projection, in addition to the dependence on $n$ and $\beta_{\text{int}}$. Above we have assumed that $\beta$ is sufficiently large for quantities computed by integration up to $\beta_{\text{int}}$ have converged to the $\beta \to \infty$ limit, and in the QMC simulations we also monitor this convergence. Naturally, the value of $\beta$ required grows with $\beta_{\text{int}}$.

### 2. Performance check for the Heisenberg chain

We have checked the validity of the estimator (37) for the Heisenberg chain of 16 spins with periodic boundary

conditions. The projection length $\beta$ is set large enough to see the asymptotic behavior of the dynamic correlations on the imaginary-time axis. The correlations were measured from the center of the QMC configuration; that is, from $\tau_0 = \beta/2$. As in the previous section, the loop algorithm and improved correlation-function estimator [5] are used. We consider the triplet gap at $k = \pi$ and the operator $A$ in Eq. (21) is then explicitly given by $A = \sum_r \mathbf{S}_r (-1)^r$ and

$$A^\dagger e^{\tau H} A = \sum_{\alpha=x,y,z} \sum_{r,r'} S_r^\alpha e^{\tau H} S_{r'}^\alpha (-1)^{r-r'}. \quad (39)$$

The Fourier-transformed correlation functions (35) were directly measured in the simulation, and no discretization error is introduced. The gap estimators (37) for several $n$ and $\beta_{\text{int}}$ were calculated and the errors estimated by the jackknife analysis [45] (bootstrapping would produce statistically equivalent result).

As shown above, our estimator (37) converges to the exact gap in the limit of infinite $n$ and $\beta_{\text{int}}$. In practice, if good convergence is observed within statistical errors, a gap value obtained from a sufficiently large $n$ and $\beta_{\text{int}}$ within the converged range can be used as the final estimation, or some appropriate functional form can be used for extrapolation. However, an important issue here is that the statistical errors grow with the two parameters. We therefore inevitably encounter a trade-off problem between systematical and statistical extrapolation errors. Here we will demonstrate that the statistical error can be optimized by proper extrapolations while keeping the estimation unbiased. In the procedure used below, the extrapolation is taken for $n \to \infty$ first and then for $\beta_{\text{int}} \to \infty$, though, as discussed above, other ways to accomplish the limit are also possible.

Examples, based on $2^{20}$ ($\sim 10^6$) Monte Carlo samples, of $n \to \infty$ extrapolations are shown in Fig. 5. The leading finite-$n$ correction is linear in $1/n$ [in accord with Eqs. (A4) and (A10) in Appendix], and we include also a quadratic term to obtain good fits to the data for the full $n$ range. Note that the errors for different $n$ are correlated; thus the curves look much more smooth than they would be with the shown error bars if the data were uncorrelated. The
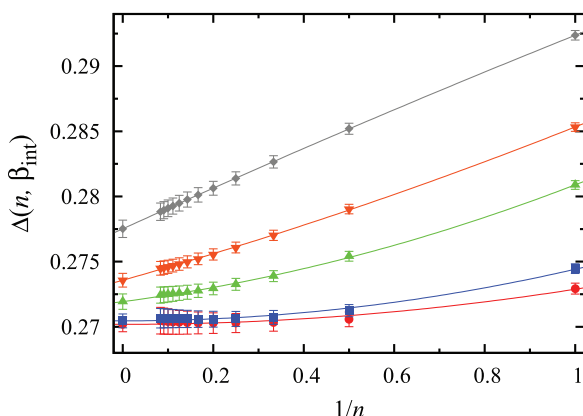


FIG. 5. (Color online) Extrapolation of $\Delta(n \to \infty, \beta_{\text{int}})$ for $\beta_{\text{int}} = 6$ (diamonds), 7.68 (lower triangles), 9.6 (upper triangles), 19.2 (squares), and 32 (circles). A quadratic polynomial in $1/n$ was fitted to the data points for each $\beta_{\text{int}}$ and the extrapolated $n \to \infty$ value with error bar is shown in each case.
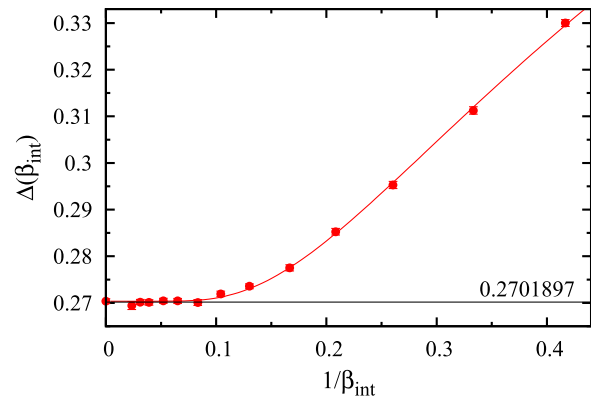


FIG. 6. (Color online) Extrapolation of the gap for $\beta_{\text{int}} \to \infty$. An exponential function $f(\beta_{\text{int}}) = \Delta + a \times \exp(-b\beta_{\text{int}})$ (curve) has been fitted to the data points, where $a$ and $b$ are positive real numbers. The exact gap value calculated from the full diagonalization is shown by the horizontal line.

minor upturn (barely visible in the figure) for small $1/n$ in the case of $\beta_{\text{int}} = 19.2$ is consistent with a statistical fluctuation; the coefficient of the linear term in $1/n$ should be positive according to the asymptotic form (A10). Here taking into account the covariance between plots would improve the result. Nevertheless, the extrapolated estimation without consideration for the covariance is adequately unbiased as we will show below. Thus we expect that the extrapolated values should not be sensitive to the covariance.

The limit $\beta_{\text{int}} \to \infty$ is taken next using the resulting values of the $n \to \infty$ extrapolations, as shown in Fig. 6. Here an exponential function is used for the data fit (again, according to results derived in the Appendix) to extrapolate the final result for the gap. Though other ways of extrapolating to $n, \beta_{\text{int}} \to \infty$ are possible, the above protocol is convenient because the extrapolation for $n$, which is taken first, is relatively easier than that for $\beta_{\text{int}}$.

As a test of the unbiased nature of the extrapolation scheme, distributions of the relative gap error are shown in Fig. 7. Histograms were collected based on 2048 independent simulations of the $L = 16$ Heisenberg chain. Results based on the extrapolation procedure discussed above are compared with those of individual gap estimators for $(n, \beta_{\text{int}}) = (1, 32)$ and $(10, 32)$. The $(1, 32)$ estimator clearly has a nonzero systematic error remaining, which is similar to (but smaller than) the conventional second moment estimator [28]. The $(10, 32)$ estimator has a small enough systematical error (the histogram being centered very close to zero) but has a large statistical error (wide distribution). The extrapolated estimation is unbiased and has a small statistical error.

### III. THE VELOCITY FROM WINDING NUMBER FLUCTUATIONS

We here discuss how to use winding numbers to compute the velocity. This method has been known for some time [12] and was recently applied to the 2D XY model [13] and the 2D Heisenberg model [6]. We begin by briefly recollecting some key aspects about winding numbers in finite-temperature QMC
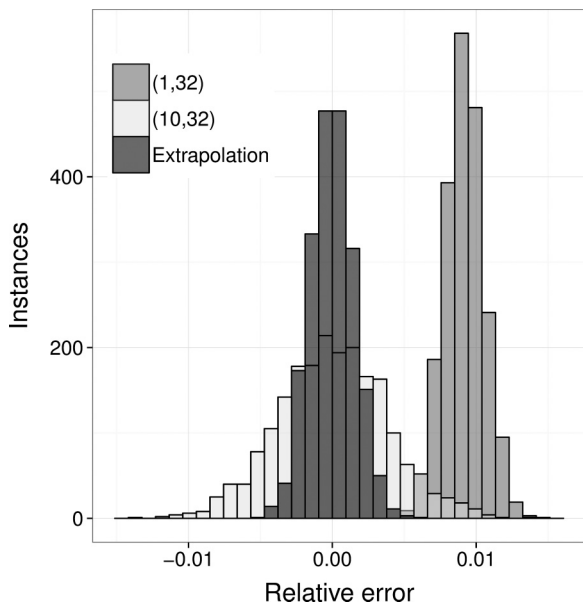
FIG. 7. Histograms of the density of the relative error of the $(n \to \infty, \beta_{\mathrm{int}} \to \infty)$ extrapolated gap, along with those based on the estimator (37) for fixed $(n, \beta_{\mathrm{int}}) = (1, 32)$ and $(10, 32)$. The error was calculated as $(\hat{\Delta} - \Delta)/\Delta$, where $\hat{\Delta}$ is the generalized-moment estimated value from $2^{20}$ ($\sim 10^6$) Monte Carlo samples and $\Delta$ is the exact value for the $L = 16$ Heisenberg chain. The histograms include results of 2048 independent simulations.

simulations, in particular the SSE method we use for these calculations. We then present results for the Heisenberg chain.

### A. Winding numbers in QMC simulations

QMC simulations at finite temperature are based on some mapping of the partition function of a quantum system in $D$ dimensions to an effectively equivalent $(D + 1)$-dimensional classical system, where the additional dimension of the configurations corresponds to imaginary (Euclidean) time. The effective length of the system in the time dimension is $c\beta$, where $\beta = 1/T$ (setting $k_B = 1$) and the configurations are time-periodic. For a system with conserved particle number, which in the case of spin model corresponds to conserved magnetization along the quantization ($z$) axis, imposing periodic boundary in the spatial dimensions leads to another, topological number associated with the configurations—the winding number representing permutations when particles are propagated once or multiple times around the periodic system. The winding numbers were first used in QMC calculations of the superfluid stiffness of interacting bosons [46].

The SSE QMC algorithm [4,32,47] is based on a Taylor expansion of the imaginary-time evolution operator (the Boltzmann operator) $e^{-\beta H}$, and similarities with the projector approach discussed in the previous section were already pointed out. Each SSE configuration is associated with some power $H^n$ of the Hamiltonian propagating a basis state (here in the standard computational basis of $z$ spin components), and these powers are sampled stochastically to all contributing orders. The trace over all basis state is also sampled. The average expansion power $\langle n \rangle$ in this procedure is proportional

to $\beta$; $\langle n \rangle = \beta \langle H \rangle \propto \beta N$. In simulations, the state propagation is broken up into individual paths corresponding to strings of $n$ of the individual local terms of the Hamiltonian, forming successions of $n$ evolving basis states, similar to those in path integrals. For a Heisenberg model, the Hamiltonian terms are the diagonal operators $S_i^z S_j^z$ (in practice with a constant added) and off-diagonal $S_i^+ S_j^- + S_i^- S_j^+$ operators, the latter of which transport spin and are associated with currents $J_a = \pm 1$ in the lattice direction $a$ corresponding to the site-pair $i, j$. The winding number in the $a$ lattice direction is defined in terms of the currents as

$$W_a = \frac{1}{L_a} \sum_{p=1}^{n} J_a(p), \tag{40}$$

where the index $p$ corresponds to the location of the transport "event" in the string of $n$ operators and $L_a$ is the length of the lattice in the $a$ direction. Defined in this way, the winding numbers are integers. It should be noted that the definition of the winding number is exactly the same in SSE and world-line methods [48,49] and one can also think of the SSE configurations as consisting of world lines (for up and down spins in the case of $S = 1/2$ quantum spin systems).

The spatial winding number $W_r$ measures the net spin transported around the periodic lattice in the $r$ direction in the course of the periodic propagation in imaginary time. Equivalently, this is the number of world lines (up ones minus down ones divided by two) crossing through a plane drawn along the time axis perpendicular to the $r$ axis. Since the total $z$ magnetization $M_z$ is conserved, one can also think of $M_z$ as a winding number; the net number of world lines crossing a plane drawn at an arbitrary time point perpendicular to the time axis, which is just the magnetization computed in the stored basis state;

$$W_\tau = M_z = \sum_{i=1}^{N} S_i^z. \tag{41}$$

The expectation values of the squared winding numbers (i.e., the winding number fluctuations) are related to two important thermodynamic quantities; the spin stiffness

$$\rho_s = \frac{1}{2\beta} \left( \langle W_x^2 \rangle + \langle W_y^2 \rangle \right) \tag{42}$$

and the uniform magnetic susceptibility

$$\chi = \frac{\beta}{N} \langle M_z^2 \rangle = \frac{\beta}{N} \langle W_\tau^2 \rangle. \tag{43}$$

The technicalities of implementing these observables in SSE calculations have been discussed extensively in the literature (see Ref. [33] for a recent review).

### B. The cubic criterion and the velocity

In the high-temperature limit $T \to \infty$, the magnetization fluctuations of any system are maximized and therefore $\langle W_\tau^2 \rangle > 0$ according to Eq. (43). For an unfrustrated antiferromagnet, the ground state is a singlet, and, on account of the presence of a singlet-triplet finite-size gap, $\langle W_\tau^2 \rangle \to 0$ when $T \to 0$ for any finite system. In contrast to these limits of $\langle W_\tau^2 \rangle$, for the spatial winding number in direction $r$ ($r = x, y, \ldots$), we

have $\langle W_r^2 \rangle \to 0$ when $T \to \infty$ on account of there being no quantum fluctuations when the imaginary-time length $\beta \to 0$ and there are no contributions from expansion powers $n > 0$ (in the case of SSE—in world-line methods there will similarly be no transport events causing shifts of the world lines). In the limit $T \to 0$, for a system with long-range order (or a "quasiordered" 1D system with power-law decaying correlations), the stiffness constant converges to a nonzero value for any $L$, and according to Eq. (42), we must then have a divergence $\langle W_r^2 \rangle \sim 1/T$. These different behaviors of the spatial and temporal winding numbers versus temperature guarantees that there is a crossing point,

$$\langle W_r^2(\beta^*) \rangle = \langle W_\tau^2(\beta^*) \rangle, \tag{44}$$

at some unique value of $\beta = \beta^*(L)$ for given size $L$.

The winding numbers characterize global fluctuations of the system in the different spatial and temporal directions. It is then natural to define a system as having cubic space-time geometry when Eq. (44) holds (with the lattice length $L$ the same in all spatial directions). The aspect ratio $L/\beta^*$ should then directly correspond to the velocity of the long-wave-length excitations. In some cases, this can be shown directly based on low-energy field theory [6,12,13], but even in the absence of such descriptions the arguments are very general and one can expect the conclusion $c = L/\beta^*$ to always hold for a system with linear dispersion, though we are not aware of any formal proofs in the general case.

### C. Test on the Heisenberg chain

To find the point where the cubic criterion $\langle W_x^2 \rangle = \langle W_\tau^2 \rangle$ is satisfied, we simulate a system at several values of $\beta$ in the region where $\langle W_x^2 \rangle \approx \langle W_\tau^2 \rangle$ based on initial explorations and knowledge of the approximate value of the velocity. We fit a low-order polynomial (typically second- or cubic-order) to the difference $\langle W_x^2 \rangle - \langle W_\tau^2 \rangle$ and solve the resulting equation for the $\beta$-value for which the cubic criterion is satisfied. This procedure is illustrated in Fig. 8 for a Heisenberg chain with $L = 64$ spins. From this procedure, we obtain $c(L) = L/\beta^*(L)$, which can be extrapolated to $L \to \infty$. With



FIG. 9. (Color online) Size dependence of the velocity obtained with the cubic criterion for the Heisenberg chain, along with a fit corresponding to a leading $\sim 1/L^4$ size correction. The dashed line is at the rigorously known velocity $c = \pi/2$.

independent data points, the statistical error of $c(L)$ at fixed $L$ can be determined by repeating the procedure multiple times with added Gaussian noise of standard deviation equal to the error bars of the data points, whence the standard deviation of the extracted crossing point is the error bar. Note that the recently proposed autotuning technique, or the stochastic approximation [50], could be used for the $\beta$ optimization instead of the interpolation of simulated data.

Figure 9 shows results of such a procedure for several chain lengths $L$, graphed versus $1/L$. We are not aware of any theoretical predictions for the size dependence of this definition of $c(L)$, but the data for the larger systems are well described by a constant (the final infinite-size value of $c$) plus a term proportional to $1/L^4$. Using this fitting form leads to a value of $c$ completely consistent with the known value $c = \pi/2$, as is clear from Fig. 9.

### IV. HYDRODYNAMIC RELATIONSHIP

A well known way to extract the velocity in an antiferromagnet is to use the analog of a hydrodynamic relationship between the velocity, the spin stiffness (helicity modulus), and the transversal susceptibility (which is the analog of a mass density) [14]:

$$c = \sqrt{\frac{\rho_s}{\chi_\perp}}. \tag{45}$$

In a QMC calculation in which the spin-rotation symmetry is not explicitly broken, one cannot compute $\chi_\perp$ directly, but one can use the fact that the rotationally-averaged uniform susceptibility $\chi$ computed in the $z$ basis, as in Eq. (43), becomes $2/3$ of a transversal (e.g., $x$) component when $T \to 0$ in the thermodynamic limit (since the longitudinal component vanishes at $T = 0$). Thus, one can obtain $\chi_\perp$ as $(3/2)\chi$ by taking the limit $L \to \infty$ before the $T \to 0$ limit, while taking the limits in the opposite order does not work because then $\chi \to 0$ due to the finite-size gap between the $M_z = 0$ and $M_z > 0$ magnetization sectors. In contrast, for $\rho_s$, the limit $T \to 0$ has to be taken before $L \to \infty$, because the system in the thermodynamic limit only has stiffness (Néel order) exactly at $T = 0$. In this case, too, a factor $3/2$ has to be included in
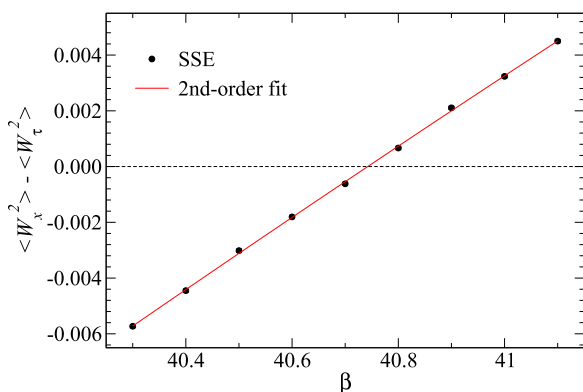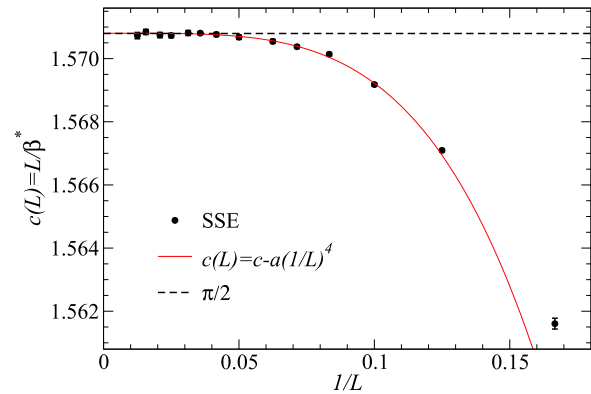


FIG. 8. (Color online) Difference between the spatial and temporal winding numbers vs the inverse temperature in simulations of a 64-site Heisenberg chain. A second-order polynomial has been fitted to the data points, and this curve is used to determine the value $\beta = \beta^*$ at which cubic condition $\langle W_x^2 \rangle = \langle W_\tau^2 \rangle$ holds (i.e., intersection with the horizontal dashed line).

the finite-size estimate to account for rotational averaging of the relevant transversal components.

The procedure to obtain $\rho_s$ in the thermodynamic limit is relatively straight-forward with an extrapolation of $\rho_s(L, T \to 0)$ using a polynomial in $1/L$, while the extrapolations requiring $L \to \infty$ first in the $\chi$ calculation is more cumbersome. Results obtained for $c$ in this way [33] have large error bars compared to the result of the winding number method presented above in Sec. V A.

In order to define a finite-size velocity estimate $c(L)$ based on Eq. (45) directly in the $T = 0$ limit we here use a modification of the relationship. We use the susceptibility at finite momentum **q**,

$$\chi(\mathbf{q}) = \frac{1}{N} \int_0^\beta \langle M_z(-\mathbf{q}, \tau) M_z(\mathbf{q}, 0) \rangle, \qquad (46)$$

where the magnetization at nonzero momentum is the same as $A$ in Eq. (7) but without the normalization by $N^{-1/2}$. We can use the smallest momentum $q = 2\pi/L$ for a given system size and approach the $q = 0$ limit as $L \to \infty$. If we here define $\rho_s(L)$ without the rotational factor $3/2$ discussed above, then the same rotational factor should also not be used in the susceptibility, and we define the velocity for finite size as

$$c(L) = \sqrt{\frac{\rho_s(L)}{\chi(q = 2\pi/L)}}. \qquad (47)$$

We will compute this quantity using SSE simulations at sufficiently large $\beta$ to achieve ground-state convergence for each $L$ studied.

**Test on the Heisenberg chain**

The hydrodynamic relationship (45) is normally applied in the magnetically ordered Néel state and one may then question its use in a critical system. In the case of the Heisenberg chain we can rigorously see that it is a valid way to extract $c$. Since there is no long-range order, there is no distinction between longitudinal and transversal modes, but Eq. (45) defined with rotationally averaged quantities should remain valid. According to the exact Bethe ansatz solution, in the thermodynamic limit we have, in the units with $J = 1$ and all relevant constants set to 1, $\rho_s = 1/4$ (derived in Ref. [51]) and $\chi(q \to 0) = 1/\pi^2$ (see., e.g., Ref. [52]) and, thus, according to Eq. (47), we obtain the correct result $c = \pi/2$.

For finite system size, it is well known that both $\chi$ and $\rho_s$ exhibit strong logarithmic corrections to their infinite-size values, which can be traced to the presence of marginal operators and it is very difficult to extrapolate them; see, e.g., Ref. [53] for $\rho_s$ and Ref. [52] for $\chi$ (where in the latter case the logarithmic correction in temperature is discussed but one should expect similar corrections for $q \to 0$ at $T = 0$). The logarithmic corrections do not cancel in the ratio of the two quantities in Eq. (47), and as a consequence we also find that it is difficult to extrapolate $c(L)$ precisely to the thermodynamic limit. Results are shown in Fig. 10. Although it is not possible to fully extrapolate to the exact results based on these data for system size up to $L = 512$, the result for the largest size nevertheless deviates by only 0.2% from the exact result.
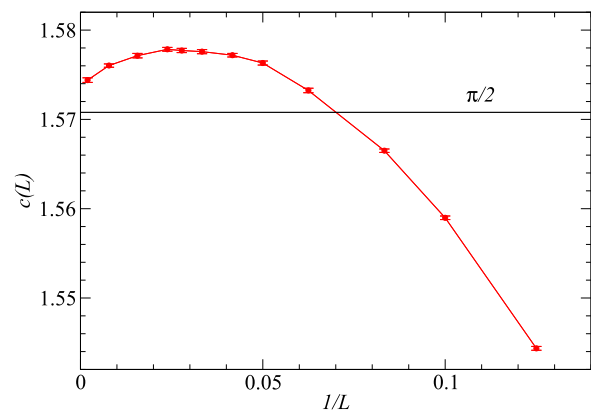


FIG. 10. (Color online) Size dependence of the velocity of excitations in the Heisenberg chain, defined according to the hydrodynamic relationship in the form (47). The apparent poor convergence to the exact result $c = \pi/2$ when $L \to \infty$ can be explained by logarithmic corrections. The maximum can be related to the anomaly in the excitation spectrum at $k = 2\pi/L$, which causes the maximum at the same $L$ in Fig. 4.

It is interesting to note that also this definition of $c(L)$ exhibits a nonmonotonic behavior, with a maximum at $L \approx 40$, the same as in the $c(L)$ value obtained previously from the gap at $k = 2\pi/L$ in Fig. 4. In the latter case, the maximum corresponds to an elevated excitation energy at $k = 2\pi/L$, which one should expect to lead to a reduction in $\chi(2\pi/L)$ because this quantity is given by a sum rule of $S(k, \omega)/\omega$. By this sum rule, in a single-mode approximation a larger $\omega_k$ would lead to a smaller $\chi(k)$, and even beyond the single-mode approximation one should expect such an effect because the dominant spectral weight is at the gap. If $\rho_s$ is not appreciably affected by this finite-size anomaly, then indeed an elevated gap and reduced $\chi(2\pi/L)$ in Eq. (47) can explain the maximum in Fig. 10.

In light of the logarithmic corrections to $c(L)$ defined according to Eq. (45), the apparent complete lack of any challenging corrections or nonmonotonic behavior in the definition based on the cubic criterion for winding numbers (Fig. 9) is even more remarkable, since also this estimate involves quantities directly related to the susceptibility (temporal winding number) and spin stiffness (spatial winding number). There are also no apparent logarithmic corrections in the velocities defined based on gaps in Figs. 3 and 4. In Fig. 3, which is based on the triplet gap close to $\pi$, the logarithmic corrections are avoided by subtracting the $k = \pi$ gap in Eq. (18), while the gap close to $k = 0$ used in Fig. 4 is not expected to be affected by logarithms. The latter gap at $k = 2\pi/L$ scales as $ck + b/L^2$ for large $L$ according to our results in Fig. 4, which can be explained from the known dispersion $\Delta_k = c \sin(k) = ck + dk^3 + \ldots$, if there is a finite-size correction $\sim 1/L^2$ (due to irrelevant fields only).

## V. 2D HEISENBERG MODEL

The 2D spin-1/2 Heisenberg model on the square lattice spontaneously breaks spin rotation symmetry at $T = 0$ in the thermodynamic limit. This leads to gapless linearly dispersing Goldstone modes in the vicinity of the momenta $(0,0)$ and

$(\pi,\pi)$. The ground state in a finite periodic system is a total spin singlet. The long-range antiferromagnetic order in the thermodynamic limit is reflected in the energies of the $S > 0$ quantum rotor states [55], which collapse onto the ground state as $\Delta_S \sim 1/L^2$ much faster than than the spin wave excitations, $\Delta \sim 1/L$. The rotor states, thus, become degenerate with the ground state as the system size increases, and have momenta $\mathbf{k} = (0,0)$ and $(\pi,\pi)$ for even and odd $S$ respectively. Combinations of rotor states with $S$ up to $\sim L$ can then be formed which are ground states with fixed direction of the Néel vector (the staggered magnetization), thus allowing for symmetry breaking in the thermodynamic limit. Here we provide an accurate determination of the spin-wave velocity and also directly investigate the scaling of the rotor state.

### A. Velocity from winding numbers

Since we expect the winding numbers with the cubic criterion to provide the best determination of $c$ we start with this approach. Results based on the procedures discussed in the preceding section are shown in Fig. 11 versus the inverse system length. Carrying out polynomial fits, we find that no linear and quadratic terms are required. Fourth- and fifth-order polynomials fits excluding these terms and using all the $L \geqslant 6$ data give almost identical results for the $L \rightarrow \infty$ extrapolated velocity, with only the error bar somewhat larger for the fifth-order fit. The figure shows the fifth-order fit and the extrapolated velocity with it is $c = 1.65847(4)$. While we do not know the physical reason for the leading cubic correction, we use it as the simplest empirical description of the data. Including also a quadratic term, it comes out equal to 0 within statistical errors and the extrapolated result does not change appreciably.

The above value of $c$ agrees within errors bars with the recent result using the same method by Jiang and Wiese [6], but our error bar is almost an order of magnitude smaller. We note that in Ref. [6] no systematic extrapolation was carried out to the thermodynamic limit—instead an average was taken

of results for system sizes in the range $L \in [24,64]$. Looking at the data in Fig. 11, it is clear that, with the small error bars on the SSE data we have achieved here, an extrapolation is necessary to obtain a result with no remaining finite-size effects. To our knowledge, the above result is the most precise spin-wave velocity reported to date for the 2D Heisenberg model. Spin-wave theory with corrections up to order $1/S^2$ gave $c = 1.6638(3)$ [54], where the uncertainty 3 in the last digit reflects estimated numerical errors from evaluations of challenging integrals. Thus, to this order, the spin-wave result deviates by only 0.3% from the correct value.

### B. Gap scaling of rotor states

The quantum rotor excitation gap can be directly accessed by measuring the lowest triplet gap at $\mathbf{k} = (\pi,\pi)$. This energy scale is related to the uniform (transverse) magnetic susceptibility $\chi_\perp$ as [55]

$$E(S,L) = \frac{S(S+1)}{2L^2\chi_\perp}. \tag{48}$$

From the behavior of the lowest triplet gap up to system sizes $L = 64$, as shown in Fig. 12, we indeed observe that $\Delta(\pi,\pi) \sim 1/L^2$ at large $L$, but there are also large corrections which we fit with additional higher-order powers of $1/L$. The extrapolation to infinite size gives $1/2\chi_\perp = 7.62(2)$. This is consistent with the value of susceptibility obtained using QMC calculations in small external magnetic fields to extract gaps between different spin sectors [56].

### C. Velocity from gaps

The velocity of the spin waves can be estimated by measuring the triplet gap in the vicinity of $(\pi,\pi)$ and $(0,0)$. We here choose to measure the triplet gap at $\mathbf{k}_1 = (\pi + 2\pi/L,\pi)$ [or, equivalently, at $(\pi,\pi + 2\pi/L)$, which we use for averaging] since the lowest triplet excitation energy is at $(\pi,\pi)$ and $\mathbf{k}_1$ is the closest allowed wave vector to $(\pi,\pi)$ for a periodic system with linear size $L$. For this model, we have carried out only $H$-power QMC simulations. We illustrate the gap extraction
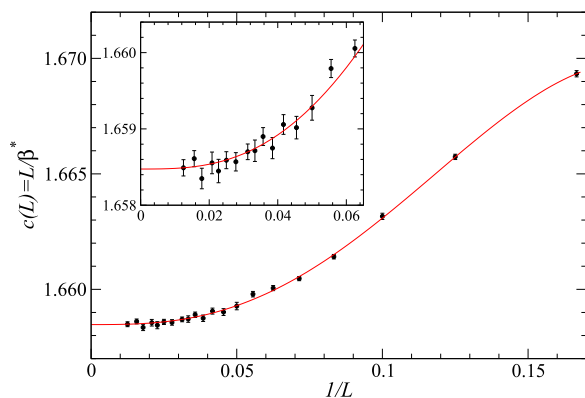


FIG. 11. (Color online) Size dependence of the velocity defined using the cubic-criterion for the 2D Heisenberg model, along with an fifth-order polynomial fit (including all system sizes $L \geqslant 6$) without linear and quadratic terms. The extrapolated velocity is $c = 1.65847(4)$. The data for the largest system sizes are shown on a more detailed scale in the inset. The goodness of the fit is $\chi^2/\mathrm{dof} \approx 0.8$.
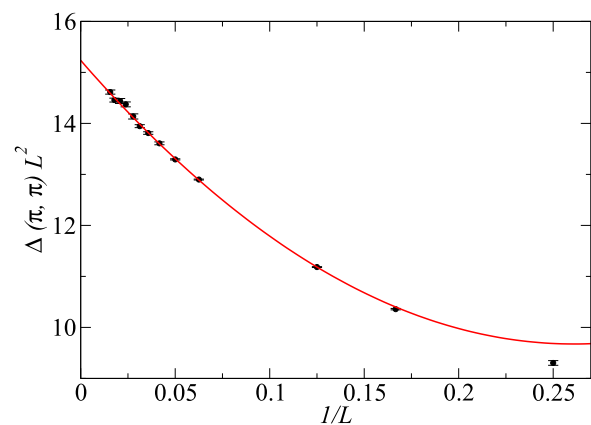


FIG. 12. (Color online) Size dependence of the lowest triplet gap $\Delta(\pi,\pi)$ multiplied by $L^2$ for the 2D Heisenberg model. This analysis shows that the gap scales as $1/L^2$ for large $L$ and an extrapolation based on a polynomial in $1/L$ gives the estimate $1/2\chi_\perp = 7.62(2)$ for the uniform susceptibility.
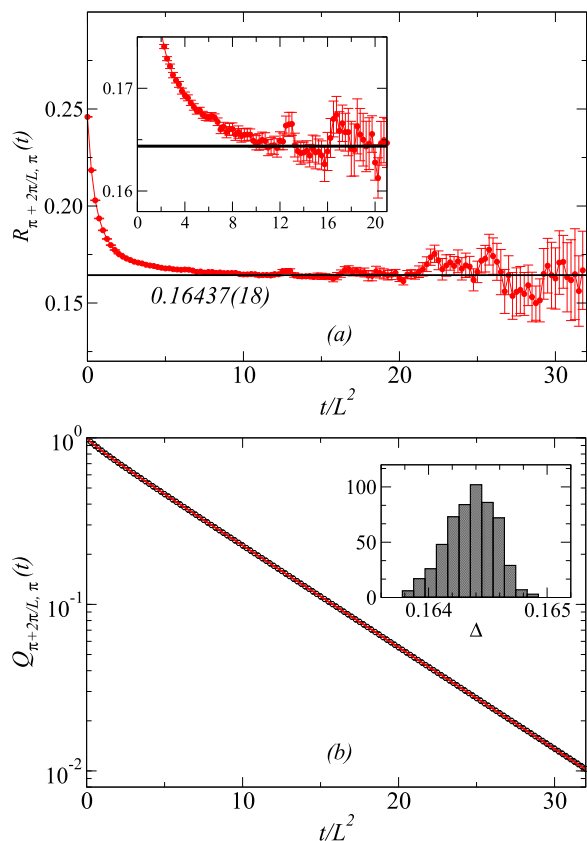
FIG. 13. (Color online) Extraction of the gap at $k = (\pi + 2\pi/L, \pi)$ based on power-projected correlation functions, shown for the largest 2D Heisenberg system studied; $L = 64$. (a) The ratio $R_k(t)$, Eq. (10) converges to a finite gap value but the error bars are large for long imaginary times. Fitting $Q_k(t)$ to four $\delta$ functions according to Eq. (11), as shown in (b), provides a more reliable gap estimation, as illustrated in the inset of (a). The thickness of the solid line here approximately represents the error bar. The inset of (b) shows the distribution of gap values obtained from a large number of different bootstrap samples of the QMC data, from which the error bar of the gap estimate was computed.

with both the ratio method and the simplified spectral function with fitted $\delta$ functions in Fig. 13, using the largest system sizes considered; $L = 64$. Again, working with the spectral function produces much more stable results, though clearly the correlation ratio $R_k$ also converges to a constant consistent with the same gap.

Since in the thermodynamic limit, the excitation energy of the spin waves equals $E(\mathbf{k}) = c|\mathbf{k} - (\pi,\pi)|$ in the vicinity of $(\pi,\pi)$, where $c$ is the spin-wave velocity, the estimator

$$c(L) = \frac{L}{2\pi}\Delta(\pi + 2\pi/L, \pi) \quad (49)$$

should converge to $c$ as $L \to \infty$. We graph this quantity in Fig. 14, and it converges to the value of $c$ obtained above from the winding number method when $L \to \infty$. Note that there is again a nonmonotonicity as a function of $L$, similar to the case of the Heisenberg chain in Fig. 4. In the latter case, this behavior only was observed in the gap extracted close to $k = 0$, however, while here we have used the spectrum close to $(\pi,\pi)$.
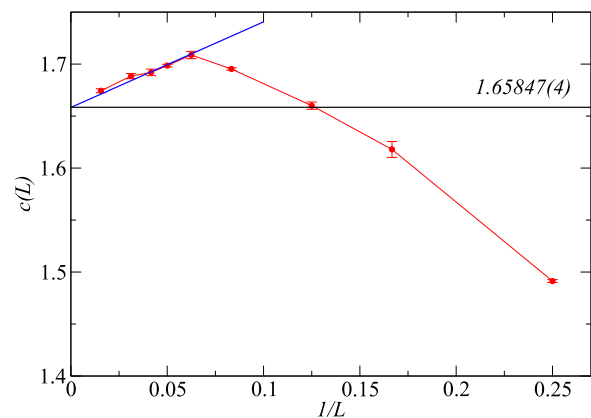


FIG. 14. (Color online) The spin-wave velocity estimator $c(L)$ defined in Eq. (49) shown for the 2D Heisenberg model as a function of $1/L$. The velocity for large $L$ is consistent with the value obtained from the winding numbers (Fig. 11), shown as a horizontal line. The line through the values for the larger system sizes is a guide to the eye.

### D. Hydrodynamic relationship

We next consider the definition of $c(L)$ by the hydrodynamic relationship, Eq. (47). We again go to sufficiently low temperature in SSE simulations for any remaining finite-$T$ corrections to be negligible compared to the statistical errors, which typically meant $\beta = 8L$. Results are graphed in Fig. 15. Here again we expect the finite size corrections in the Néel state to be described asymptotically by a polynomial in $1/L$, but a rather high order of the polynomial is required to fit the data for the smaller system sizes. The behavior for the largest sizes again indicate (as in the case of the winding number calculation discussed above) that the leading correction in $1/L$ is cubic. To get a statistically acceptable fit for all $L \geq 8$ data, we include
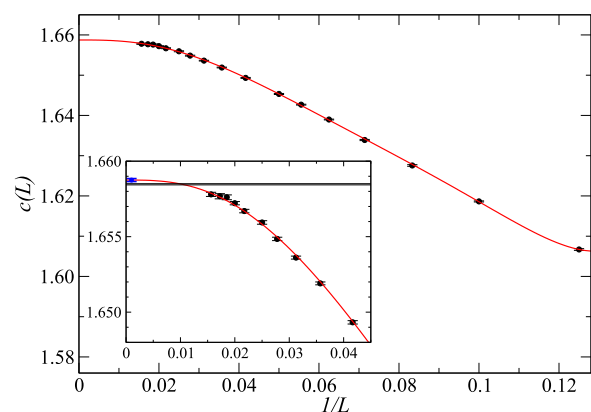


FIG. 15. (Color online) Finite-size spin-wave velocity defined using the modified hydrodynamic relationship (47). The extrapolation to infinite size is done using a sixth-order polynomial in $1/L$ without linear and quadratic terms (solid curve). The inset shows the data for the largest systems on a more detailed scale. The point (blue) with error bar close to the $y$ axis indicates the statistical error of the $L \to \infty$ extrapolation. The barely separated horizontal lines shows the value of $c$ plus-minus one error bar obtained in Sec. V A using the winding number method.

terms up to order 6, which gives $c = 1.65875(10)$, which deviates by approximately 2.5 error bars from the statistically much more precise value obtained in Sec. V A based on the winding number fluctuations. By comparing Figs. 11 and 15, it is clear that the analysis in the former is more reliable, with a larger number of data points used and an overall much weaker size dependence. The above result from the hydrodynamic relationship is still likely somewhat affected by systematical errors, as the behavior still appears to be flattening out more than the fit suggests, and the fit even including the 6th-order term is only marginally good, with $\chi^2/\mathrm{dof} \approx 1.7$. It would clearly be desirable to go to larger system sizes, but given the much better behavior of the winding number data it is already clear that this is the preferred method.

## VI. BILAYER HEISENBERG MODEL

We now consider the $S = 1/2$ bilayer Heisenberg model which is a prototypical system to study quantum phase transitions in 2D [7,8,10]. The Hamiltonian is given by

$$H = J_1 \sum_{\langle i,j \rangle}(\mathbf{S}_{i,1} \cdot \mathbf{S}_{j,1} + \mathbf{S}_{i,2} \cdot \mathbf{S}_{j,2}) + J_2 \sum_{i=1}^{L^2} \mathbf{S}_{i,1} \cdot \mathbf{S}_{i,2}, \quad (50)$$

where $\mathbf{S}_{i,a}$ represents a $S = 1/2$ spin operator at site $i$ of layer $a$ $(a = 1,2)$, and $\langle i,j \rangle$ denotes a pair of nearest neighbor sites on the square lattice of $L \times L$ sites with periodic boundary conditions. Both the couplings $J_1, J_2$ are positive (antiferromagnetic). As the ratio $g = J_2/J_1$ is increased, there is a destruction of long-range Néel order at a critical $g_c$, beyond which the system enters a disordered state with no broken symmetries. The best value available for the location of the critical point is $g_c = 2.5220(1)$ [22] and we will use this value below.

### A. Velocity from winding numbers

Results for the velocity based on the cubic winding number criterion at the critical point of the bilayer are shown in Fig. 16. Here we find that a noninteger power-law correction describes the data very well for system sizes $L \geqslant 6$. A fit of the form
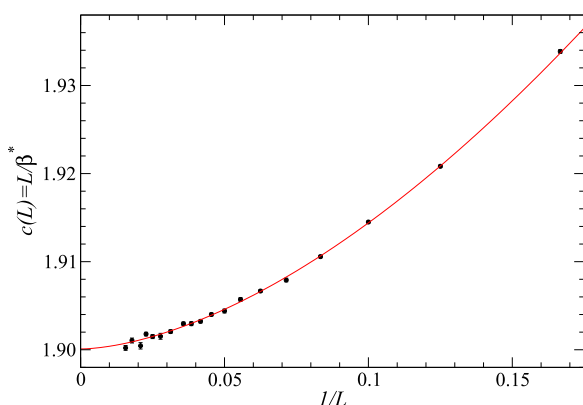


FIG. 16. (Color online) Size dependence of the velocity computed using the winding-number cubic criterion for the bilayer Heisenberg model at the estimated critical coupling ratio $g = 2.5220$. The curve shows a least-squares fit of the form $c(L) = c + b/L^a$ to the $L \geqslant 8$ data, which gives the extrapolated velocity $c = 1.9001(2)$.

$c(L) = c + b/L^a$ to the $L \geqslant 8$ data gives $c = 1.9001(2)$. Note that the fitted curve also goes through the $L = 6$ data point even though this point was not included in the fit. This adds to our confidence of this power-law correction. The value of $c$ is a few percent smaller than the spin-wave result [10] $c_{\mathrm{sw}} = 1.96$ including $1/S$ corrections at $g = 2.51$ and $c_{\mathrm{sw}} \approx 2.03$ obtained [9] to order $1/S^2$. A more sophisticated treatment beyond conventional spin-wave theory, including the effects of longitudinal fluctuations close to the critical point, gives $c = 1.78$ when the expression $c = 0.705 \times g$ below Eq. (10) of Ref. [9] is evaluated with $g_c = 2.522$. While these analytical values may appear reasonably close to (deviating by a few percent from) the presumably correct numerical value we have obtained here, the deviations are still much larger than in the case of the ordered 2D Heisenberg model, where the error of spin-wave result is only 0.3%, as discussed in Sec. V A.

The exponent of the correction in the fit in Fig. 16 is $a = 1.67(4)$. It is not clear to us how this exponent relates to the standard critical exponents of the 3D O(3) universality class of the transition, but the relatively large value (larger than $1/\nu \approx 1.41$) suggests that it may involve subleading exponents.

### B. Velocity from gaps

Since the bilayer has two sites per unit cell ($a = 1,2$ for each square-lattice point $i$), the triplet excitations have an extra label $k_z$ in **k** space which denote the in-phase ($k_z = 0$) and out-of-phase ($k_z = \pi$) spin excitations of the two layers, respectively. In the magnetically ordered Néel phase, the triplet excitations are gapless at $(0,0,0)$ and $(\pi,\pi,\pi)$, with the lowest excitation being at $(\pi,\pi,\pi)$ for a finite system. The spectrum is linear in the neighborhood of both the gapless points, which defines the corresponding spin wave velocity $c$. For a continuous phase transition, the spin wave velocity $c$ scales as [16]

$$c \propto (g - g_c)^{\nu(z-1)}, \quad (51)$$

where $z$ is the dynamical exponent and $\nu$ is the correlation length exponent. Thus, for a $z = 1$ transition as is the case for the Heisenberg bilayer at $g = g_c$, the velocity is regular at the critical point.

#### 1. Critical point

For measuring the velocity of the critical modes, we first study the triplet gap at $(\pi + 2\pi/L, \pi, \pi)$. From the linearly of the spectrum at the critical point, we define the velocity estimator in analogy with the single-layer case (49) as

$$c(L) = \Delta(\pi + 2\pi/L, \pi, \pi)L/(2\pi). \quad (52)$$

The behavior of this quantity as a function of $L$ is shown in Fig. 17. Unlike the case of the Heisenberg chain (Fig. 4) and the single layer (Fig. 14), the velocity estimate for the Heisenberg bilayer at criticality is notably higher than winding numbers estimate, by about 5% as $L \to \infty$. To extract the gap needed in Eq. (52), we have used both the power-projection QMC method with $\delta$ function fits to the correlation function, as described in Sec. II B, and the generalized moment method applied to exponential-projector QMC data, as detailed in Sec. II B 1. As seen in in Fig. 17, the two methods give results that agree fully within statistical error.
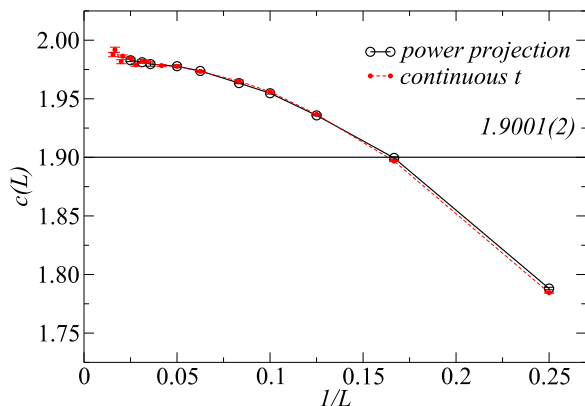
FIG. 17. (Color online) The velocity estimator (52) for the Heisenberg bilayer at its critical point shown as a function of $1/L$. Results based on both power-projection and exponential projection (continuous time) are shown and agree within error bars. The velocity estimate obtained from the winding numbers (Fig. 16) is shown as the horizontal line.

At first sight, potential reason for the disagreement with the winding number result could be an overestimation of the gap extracted from the correlation function at the critical point. Consider the full dynamic spin structure factor defined as

$$S_{\mathbf{k}}(\omega) = \sum_m |\langle m|S^z(\mathbf{k})|0\rangle|^2 \delta(\omega + E_0 - E_m). \quad (53)$$

In the Néel phase, the structure factor has a $\delta$ function of weight $A_d(\mathbf{k})$ at an energy $\omega(\mathbf{k})$ which represents the magnon mode, as well as a continuum $A_c(\mathbf{k},\omega)$, which does not extend below $\omega(\mathbf{k})$ and which decays rapidly to zero as $\omega \to \infty$:

$$S_{\mathbf{k}}(\omega) = A_d(\mathbf{k})\delta(\omega - \omega(\mathbf{k})) + A_c(\mathbf{k},\omega). \quad (54)$$

The velocity derived from $\omega(\mathbf{k})$ in the vicinity of $(\pi,\pi,\pi)$ is a regular function of $g - g_c$ and is by definition the correct velocity at the critical point. However, the weight in the $\delta$ function $A_d(\mathbf{k})$ (that represents the magnon mode) also smoothly goes to zero as the critical point is approached, with the spectrum evolving into purely a continuum reflecting the overdamped critical magnons. In our fitting scheme [see Eqs. (11) and (14)] used with the power-projection QMC method, where the spectral function is represented by a small number of $\delta$ functions, $\omega(\mathbf{k})$ may be *overestimated*, especially for large system sizes, when $A_d$ is very small close to the critical point. The velocity would then also be over-estimated, as it is in Fig. 17 (assuming that the winding number result is correct). A similar distorting effect of the continuum may be expected also with the generalized moment method used with the exponential-projection QMC data. However, the $\delta$-function fits are stable with respect to the number of $\delta$ functions used, and the extrapolations used with the generalized moment method are also stable. We do not see any evidence of remaining effects that could account for an overestimation as large as in Fig. 17. The perfect match between the two methods within their statistical errors also gives us confidence that the gaps are determined correctly and the reason for the disagreement with the winding number method must be sought elsewhere.

To search for a possible flaw in the velocity estimation based on Eq. (52), we next analyze the details of the dispersion relation around $(\pi,\pi,\pi)$, using gaps extracted with the generalized moment method. The dispersion should be an asymptotically linear function of the wave number, $\Delta(\mathbf{q}) = cq$, where $\mathbf{q}$ is the momentum relative to the gapless point $(\pi,\pi,\pi)$;

$$\mathbf{q} = \mathbf{k} - (\pi,\pi,\pi). \quad (55)$$

We should then have $\lim_{q\to 0}\lim_{L\to\infty}\Delta(\mathbf{q},L)/q = c$. There could potentially be an issue with the order of the limits $q \to 0$ and $L \to \infty$, which with Eq. (52) are taken simultaneously as we use $q = 2\pi/L$.

To investigate the formally correct limit of taking $L \to \infty$ first and then $q \to 0$, we analyze the gaps at fixed $q$ and varying $L$, to find the corresponding gap values as a function of the momentum in the thermodynamic limit. The $L$ dependence of the gaps at several different $q$-values are shown in Fig. 18. We observe that the size correction to the gap for sufficiently large $L$ is linear in $1/L$ and therefore extrapolate the data to infinite size using simple line fits. Using these gaps $\Delta(q)$, we finally extrapolate the velocity as $c = \lim_{k\to 0}\Delta(q)/q$, as shown in Fig. 19. The behavior is fully consistent with a linear dispersion in the limit $q \to 0$, and empirically we find that the leading correction is of order $q^2$ (i.e., the correction to the gap $\Delta \sim q$ is of order $q^3$). The final result for the velocity extracted this way is $c = 1.899(2)$, which is fully consistent with the result based on the winding numbers. Thus the reason for the previous disagreement is indeed that Eq. (52) does not represent the correct order of limits, and there are no flaws in the extraction of the gaps.

This more detailed analysis also shows the reason for the over-estimation of the velocity based on Eq. (52). As is clear from Fig. 18, the gaps converge to their infinite size values as $1/L$ for sufficiently large $L$. We observe that the gaps at fixed momentum $k$, where $k$ is close to but not equal to $(\pi,\pi,\pi)$,
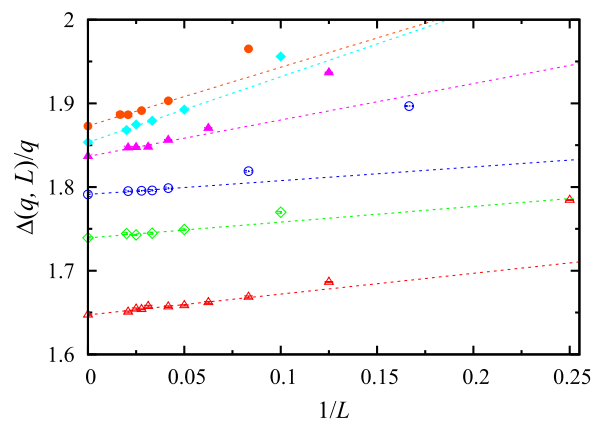


FIG. 18. (Color online) Momentum dependent gap, with $\mathbf{q}$ defined relative to $(\pi,\pi,\pi)$, divided by the respective $q$. The data are shown vs $1/L$ for $q = 2\pi/4$ (open triangles), $2\pi/5$ (open diamonds), $2\pi/6$ (open circles), $2\pi/8$ (solid triangles), $2\pi/10$ (solid diamonds), and $2\pi/12$ (solid circles). Linear fits in $1/L$ are shown for each momentum, with only data for sufficiently large $L$ included for each $q$.
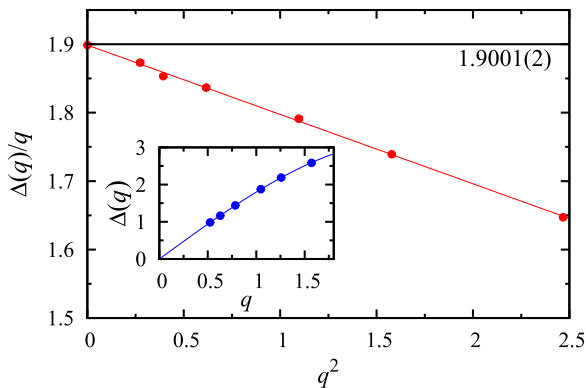
FIG. 19. (Color online) Extrapolation of the critical bilayer velocity using the gaps for $q \to 0$. The infinite-size velocity is estimated as the limiting value $c = 1.899(2)$, which is consistent with the estimation from the winding numbers, $c = 1.9001(2)$, indicated by the horizontal line. A linear function in $1/q^2$ is used for the extrapolation. The inset shows the calculated dispersion.

follows the following behavior for large $L$:

$$\Delta(q,L) = \Delta(q,L \to \infty) + \frac{A(q)}{L^2} \quad \text{(gapless)},$$

$$\Delta(q,L) = \Delta(q,L \to \infty) + \frac{B(q)}{L} \quad \text{(critical)},$$

$$\Delta(q,L) = \Delta(q,L \to \infty) + a(q)e^{-b(q)L} \quad \text{(gapped)}.$$

In the gapless and the gapped phases, the size correction decays sufficiently rapidly, so that $\Delta(q,L \to \infty) = ck$. However, at the critical point, the estimator $\Delta(q,L)L/2\pi$ instead converges to $c + B(q \to 0)/2\pi$.

### 2. Paramagnetic phase

For completeness we next extract the velocity of the propagating triplet excitations, or triplons, also in the quantum-disordered phase of the bilayer. We choose two points well away from $g = g_c$, at $g = 3$ and $4$. The lowest triplet gap at $(\pi,\pi,\pi)$ now converges to a finite ($g$ dependent) value as $L \to \infty$, because the paramagnetic phase is gapped. In the vicinity of $(\pi,\pi,\pi)$, the spectrum is expected to behave generically as

$$\Delta(\mathbf{q}) = \sqrt{\Delta_0^2 + c^2 q^2} \approx \Delta_0 + \frac{c^2 q^2}{2\Delta_0}, \tag{56}$$

where $\Delta_0$ denotes the triplet gap at wave vector $(\pi,\pi,\pi)$, $c$ is the velocity of the gapped triplons, and $q$ is again measured relative to $(\pi,\pi,\pi)$. We use both the approaches (ways of taking $q \to 0$ and $L \to \infty$) discussed above to estimate the triplon velocity $c$ and they give consistent results in the gapped phase (see Fig. 20 for the analysis at $g = 3$). We obtain $c = 1.973(4)$ and $2.159(6)$ for $g = 3$ and $4$, respectively.

The simple estimator (52) should give the correct velocity in the gapped phase, as discussed above, and we further check for consistency of the approach by also analyzing the gap at $q = 4\pi/L$, in addition to the smallest momentum $q = 2\pi/L$ (both extrapolations shown in Fig. 20, upper panel). The same velocity is also obtained using these extrapolated gap values
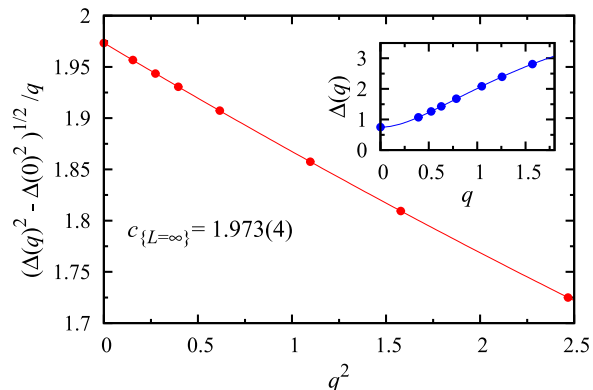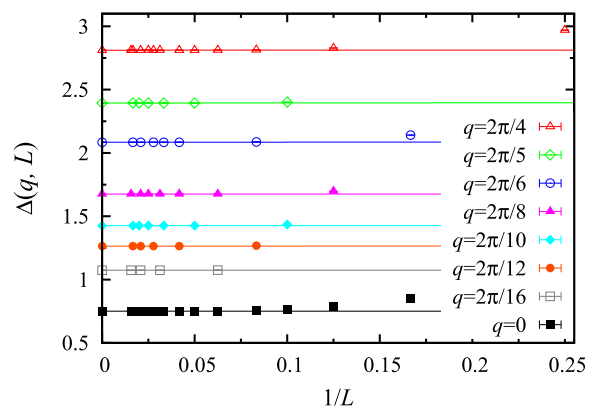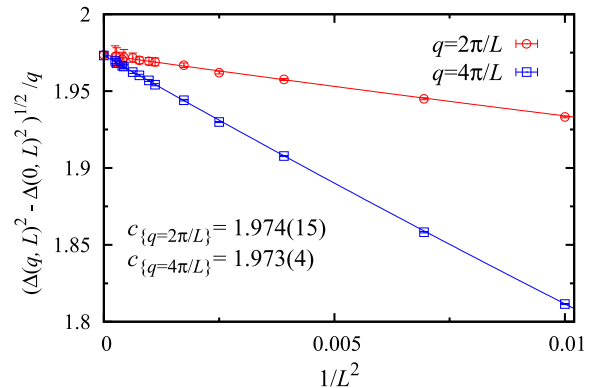


FIG. 20. (Color online) Velocity extraction for the bilayer at $g = 3$ (paramagnetic phase). The top panel shows the size convergence of velocity estimates based on two momenta close to the antiferromagnetic momentum ($q$ being the deviation from this momentum). The middle panel shows the system-size dependence of the gaps at several values of $q$. The bottom panel shows the infinite-size extrapolated velocity defined vs $q$, obtained from the fits in the middle panel and graphed vs $q^2$. The $q \to 0$ extrapolation by a polynomial (red curve) agrees with the different order of the limits taken in the top panel. The inset shows the dispersion relation.

representing the limit $L \to \infty$ before $q \to 0$, as shown in Fig. 20, bottom panel.

For completeness, we also show the velocity estimate $c(L)$ obtained from the winding numbers with the cubic criterion in Fig. 21. $c(L)$ converges (exponentially fast for large $L$) to a finite value as $L \to \infty$. However, this estimate gives an incorrect (higher) velocity in the paramagnetic phase. This is expected as the triplons are not linearly dispersing excitations
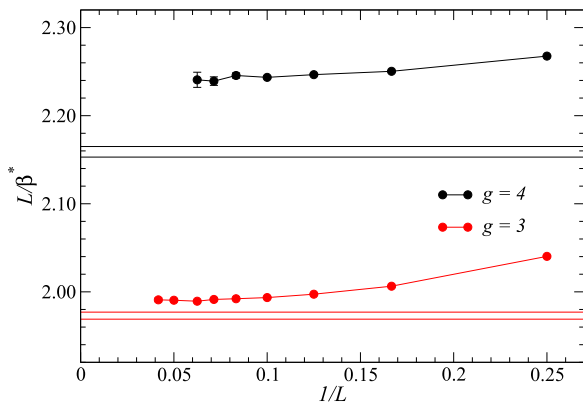
FIG. 21. (Color online) The behavior of the inverse temperature $\beta^*$ at which the cubic criterion is satisfied for the bilayer at two couplings in the paramagnetic regime. The method does not produce the correct velocity when $g > g_c$. The values, $\pm$ one error bar, obtained using the gap method are shown with the double horizontal lines.

[see Eq. (56)]. The error in the estimate increases with the distance of $g$ from the critical point, which is because $c$ is a regular function of $(g - g_c)$ from both sides and the winding number estimator does work at the critical point.

## VII. QUANTUM-CRITICAL SCALING AT $T > 0$

With $c$ of the critical bilayer determined to high precision, we have an opportunity to test in detail quantum-critical scaling predictions based on large-$N$ calculations for the O($N$) model, which for $N = 3$ should describe the universal critical behavior [15,16]. The $T = 0$ quantum-critical point influences the behavior of the system in a wide "fan" in the $(g, T)$ plane, with cross-overs to different low-$T$ behaviors away from $g_c$ set by the spin stiffness (for $g < g_c$) and the spin gap (for $g > g_c$). Here, we will test results for the temperature dependence of the uniform magnetic susceptibility $\chi(T)$ (per unit cell)

$$\chi = \frac{1}{L^2}\left\langle\left(\sum_{j=1}^{2}\sum_{i=1}^{L^2} S_{i,j}^z\right)^2\right\rangle \tag{57}$$

and the specific heat

$$C = \frac{d}{dt}E(T), \tag{58}$$

where the internal energy $E = \langle H\rangle/L^2$ per unit cell is computed with the bilayer Hamiltonian (50). In the SSE method, each configuration has a fixed $z$ magnetization (since this is a conserved quantity) and $\chi$ is evaluated directly according to Eq. (57). $C$ can be computed using an exact estimator based on the fluctuations of the number of operators in the sampled operator strings. In practice, however, this estimator is very noisy at low temperatures and we will therefore instead analyze the energy, which is simply related to the average number of SSE operators.

### A. Comparison with large-$N$ calculations

The large-$N$ treatment of the (2+1)-dimensional O($N$) nonlinear sigma-model gave the following leading-order low-$T$ forms of the susceptibility and the specific heat exactly at the critical coupling [16]:

$$\chi'(T) = \frac{\lambda}{\pi c^2}T\left(1 - \frac{0.6189}{N}\right), \tag{59}$$

where the constant $\lambda = 1.07602$, and

$$C'(T) = \frac{3\zeta(3)}{\pi c^2}T^2\left(\frac{4N}{5} - 0.3344\right), \tag{60}$$

where $\zeta(3) = 1.20206$. Since we in practice compute and analyze the internal energy, we will compare with the $T$-integrated version of (60),

$$E'(T) = E_0 + \frac{\zeta(3)}{\pi c^2}T^3\left(\frac{4N}{5} - 0.3344\right). \tag{61}$$

We use primes on the symbols above to indicate that these are not expected to be exact forms. Several tests have been reported in the literature for different variants of dimerized Heisenberg models and the leading power laws above have been confirmed. However, quantitatively, the degree of agreement has not been that well established, because $c$ has also typically been extracted from quantities relying on the large-$N$ expansions, instead of using an independently determined unbiased value. To avoid the dependence on $c$, one can also consider the Wilson ratio,

$$W = \frac{\chi T}{C}, \tag{62}$$

where $c^2$ in Eqs. (59) and (60) cancels out. Our main purpose here is to test the quantities where $c$ appears.

### B. Susceptibility

We need the susceptibility per unit cell in the thermodynamic limit. The finite-size effects are illustrated in Fig. 22 for a coupling ratio close to the quantum-critical value. To
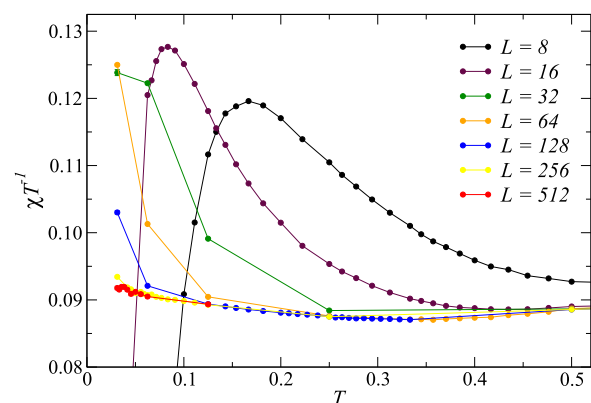


FIG. 22. (Color online) Susceptibility per unit cell divided by the temperature of the bilayer at $g = 2.5220$. Results for several different system sizes are shown to illustrate the finite size effects. For any finite $L$, the susceptibility vanishes when $T \to 0$, at a temperature scale set by the finite size gap $\Delta$, which at criticality scales as $\Delta \sim 1/L$. The lines connecting points are guides to the eye.
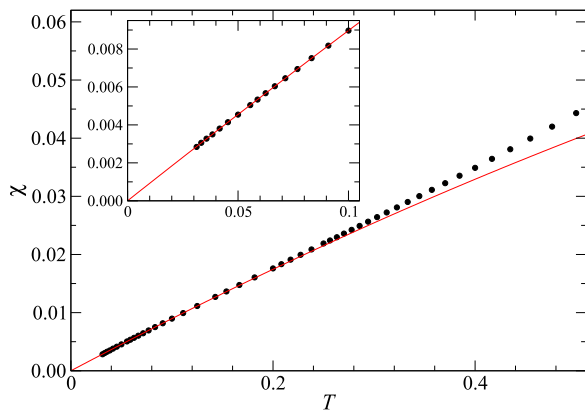
FIG. 23. (Color online) The susceptibility per spin of the bilayer Heisenberg model at coupling ratio $s = 2.5221$. The points are SSE results (with error bars much smaller than the symbols) and the curves are second-order polynomial fits to the data for $T/J_1 \leqslant 0.1$.
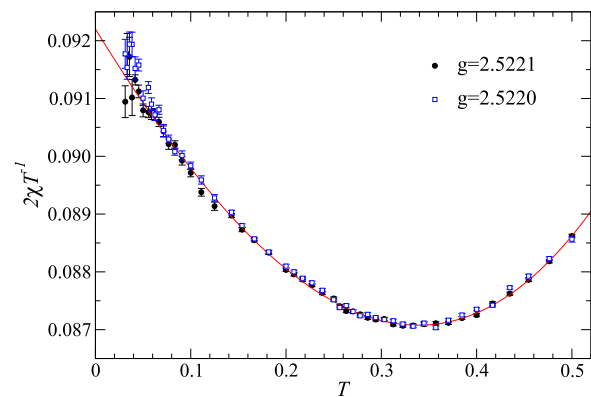


FIG. 24. (Color online) Twice the susceptibility divided by the temperature of the bilayer Heisenberg model at two coupling ratios close to the quantum-critical point. The solid curve is a third-order polynomial fit to the $g = 2.5221$ (which we estimate is closer to $g_c$) data.

make the graph clearer we have graphed $\chi/T$, which also is appropriate considering that we are here aiming to extract the size of the $T$-linear term in $\chi$. Since the magnetization is conserved and there is a finite size singlet-triplet gap $\Delta \sim 1/L$, the susceptibility for a given finite $L$ vanishes exponentially at a temperature scale $T \sim 1/L$. Interestingly, $\chi/T$ exhibits a prominent peak preceding the eventual drop to zero. The results indicate that we can compute the susceptibility without noticeable (within statistical errors) finite-size effects down to $T/J_1 = 1/32$ by using $L = 512$ lattices at the lowest temperatures (while smaller systems can be used at higher temperatures).

Below we show results for a range of temperatures representing the thermodynamic limit. In addition to calculations at $g = 2.5220$, we have also considered couplings one standard deviation of the $g_c$ estimate away from this point, i.e., $g = 2.5219$ and $2.5221$. Based on these calculations we observe that the critical point should be very close to $2.5221$. In Fig. 23, we analyze the $T$ dependence of the susceptibility. At $g = 2.5221$ a second-order polynomial fit works well with data for $T/J_1 \leqslant 0.1$ ($\chi^2/\mathrm{dof} \approx 1.2$ with 16 data points) and gives zero intercept within the statistical error of the fit, $0.000004(16)$. The slope $a$ in the expected leading-order form $\chi = aT$ is $a = 0.0922(3)$. If the intercept is assumed to be 0 and the form $\chi = aT + bT^2$ is used, the resulting slope is $0.09224(9)$. A second-order fit to data at at $g = 2.5220$ and the same range of temperatures (not shown in the figure) gives a positive intercept about two error bars away from 0, $0.000032(14)$. The slope is $a = 0.0916(6)$, statistically indistinguishable from the result at $g = 2.5221$. Although the differences between these data sets are small, they, along with results for $g = 2.5219$, suggest that the critical point is closer to $g = 2.5221$ than to $g = 2.5220$, which is still in agreement within error bars with the result $g = 2.5220(1)$ stated in Ref. [27]. Based on the present susceptibility results we estimate $g = 2.52210(5)$.

Figure 24 shows the susceptibility per unit cell with the temperature divided out at $g = 2.5220$ and $2.5221$, along with a fit to the data at the latter coupling. A third-order polynomial describe all the data well for all temperatures up to about $T/J_1 = 0.5$. The intercept is completely consistent with the results for the slope obtained in Fig. 23: $a = 0.09220(5)$. Given that this

fit includes the largest amount of data we take the result as our final estimate of the susceptibility prefactor. We can now compare it with the predicted value from Eq. (59). For $N \rightarrow \infty$ and the value of $c$ extracted in Sec. VI, we have $a' = 0.09487(2)$, i.e., 2.9% higher than our QMC estimate, which must be considered a remarkably good agreement. However, when including also the $1/N$ correction in Eq. (59), evaluated at $N = 3$, we have $a' = 0.07530(2)$, which is almost 20% too low.

### C. Specific heat

The internal energy is graphed in Fig. 25. The form $E = E_0 + bT^3$ describes well the data for $T \leqslant 2$ and a fit gives $E = -2.253040(1)$ and $b = 0.2462(6)$. The predicted factor from Eq. (61) when only the term $\propto N$ is included is $b' = 0.25435(5)$, which is 3.3% higher than our estimate. Including also the $N$-independent contribution (which here acts as the leading correction to the $N = \infty$ result) gives
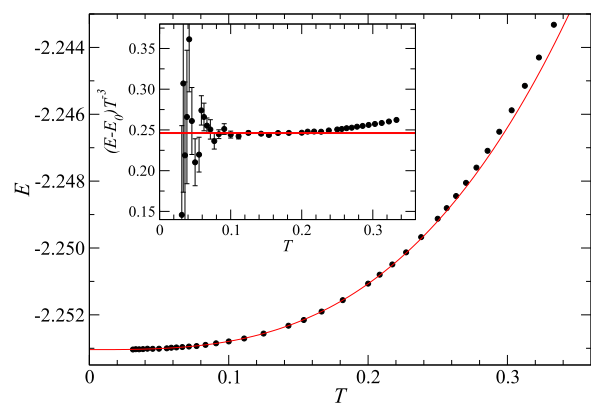


FIG. 25. (Color online) The internal energy of the bilayer Heisenberg model at coupling ratio $g = 5.5221$ (within the error bars of the estimated $g_c$), along with a fit of the form $E = E_0 + aT^3$ to the $T \leqslant 0.2$ data. The inset shows the data after subtraction of the ground-state energy and dividing by $T^3$. The horizontal line shows the prefactor of the cubic correction (with the line thickness corresponding approximately to $\pm$ one error bar) from the fit in the main graph.

$b' = 0.21891(5)$, which is about 11% too low. Thus, as in the case of the susceptibility, the $1/N$ correction makes the agreement worse.

### D. Wilson ratio

From the above expressions (59) and (60) for $\chi$ and $C$, one obtains $W' = 5\lambda/36\zeta(3) \approx 0.1243$ when only the leading terms in the $1/N$ expansion are used and $W' \approx 0.1147$ when the corrections are included. A value differing from the leading-order prediction by only about 2% was reported in Ref. [27] based on large-scale studies of a single-plane model with columnar dimerization; $W = 0.1262(6)$. The Wilson ratio with the exact prefactors $a$ and $b$ in the asymptotic forms $\chi = aT$ and $E = E_0 + bT^3$ ($C = 3bT^2$) is $W = a/3b$. With the values of $a$ and $b$ determined above we obtain $W = 0.1248(3)$. This value agrees reasonably well with the less accurate (with larger error bar) value obtained in Ref. [27], being smaller by slightly less than two combined error bars. It is in remarkably good agreement with leading value of the $1/N$ expansion quoted above, deviating by only one error bar. This good agreement again deteriorates when the leading corrections to $\chi$ and $C$ are used in $W'$.

### VIII. SUMMARY AND DISCUSSION

We have here presented several methods based on QMC simulations to extract the velocity of excitations in $S = 1/2$ quantum spin models: the one-dimensional and two-dimensional antiferromagnetic Heisenberg models as well as the bilayer Heisenberg model. The methods are fundamentally not new, but we presented several technical solutions to improve their practical utility and carried out extensive tests with the goal of testing their precision in practice.

The method of computing the lowest excitation energy for given momentum using the long-time behavior of imaginary-time correlations directly probes the dispersion relation and is, thus, directly connected to the definition of the velocity. It can be used in an unbiased manner to estimate the velocity both in the ordered and disordered phases and at the critical point. This is, however, also the most complicated method in practice, as it requires significant efforts to obtain high-precision results for appropriate correlation functions and to analyze them, trying to avoid contamination by excited states in the gap estimation. We here employed two different methods to compute the correlations, using the power and the exponential-form projection. We also presented two different ways to extract the gap, using fitting procedures sensitive to the slowest decaying component of the correlation functions and the generalized moment method [28] extended to projector QMC simulations. The good agreement of the velocity obtained in this way with the other approaches in the limit of large system sizes show that the gap estimators are unbiased (and we also demonstrated this directly for small system sizes where exact results can be obtained). We expect these gap-extraction procedures to be useful also for many other systems and studies, and to be the most versatile method between the three approaches that we investigated in the present paper.

The velocity estimator based on the gap at the momentum $\mathbf{k}$ closest, but not equal, to the momentum $\mathbf{k}_0$ of the lowest triplet excitation in the finite system [where $\mathbf{k}_0 = (\pi, \pi, \dots)$

for these SU(2) symmetric models], requires calculating the lowest triplet gap only for a single momentum [here $\mathbf{k} = (\pi - 2\pi/L, \pi, \dots)$] at each $L$. This method works extremely well deep in the disordered phase (see Fig. 20, top panel). However, this estimator does not work at the critical point as shown here (Fig. 17). In the gapless ordered phase, we see a nonmonotonic scaling of this estimator in $1/L$ (Fig. 14), which is interestingly similar to the 1D case (Fig. 4), so care must be taken to reach sufficiently big system sizes before a careful extrapolation to $L \to \infty$ may be performed. A better, though more numerically intensive way to calculate the velocity both for the gapless phase and the critical point, is to calculate the lowest triplet gap at a few fixed momenta close to $\mathbf{k}_0$ and extrapolate each gap to its $L \to \infty$ value, and then to finally fit these to the known dispersion relation. The finite-size scaling of the gaps is monotonic in $1/L$ (Fig. 18) and the method gives the correct velocity in irrespective of the nature of the phase (Fig. 19).

The method of using winding number fluctuations to define a cubic space-time geometry provides an altogether different approach to calculate the velocity and fundamentally relies on the low-energy physics of systems with dynamic exponent $z = 1$, apart from the obvious fact that the winding number exists, which imples at least O(2) spin-rotation symmetry of the system. The method requires calculations for several temperatures in the neighborhood of the target temperature $T^*$ at which the temporal and spatial winding-number fluctuations are equal, or, in principle, one could use a reweighting technique and only do a single simulation very close to the estimated crossing point, though we expect that it should be statistically advantageous to do several independent runs as more independent data are collected and the statistical analysis (curve fitting and error estimate) is much simpler. Though doing several simulations may seem like a drawback, in practice the procedure for extracting the crossing temperature are very straightforward and based on our work presented here this is the preferred method to extract the velocity. As noted in a previous application of this method [6], the convergence with the system size is very rapid, but here we still found it necessary to use a final extrapolation using a power-law correction (with integer or nontrivial leading power in a long-range ordered and critical system, respectively), to avoid any remaining size effects. This method gives the correct velocity both in the gapless phase (Fig. 11) and at the critical point (Figs. 9 and 16), but fails to work in the gapped phase (Fig. 21). It is also considerably simpler to apply computationally compared to the gap extraction method and gives the best estimate in the cases where it is applicable.

We also presented a modification of the standard hydrodynamic relationship between the velocity, spin stiffness, and susceptibility, computing the latter at nonzero momentum $q = 2\pi/L$ in order to obtain a useful finite-size estimate of the velocity (with the standard relationship at $q = 0$ diverging when $T \to 0$ in a finite system in QMC simulations). This method also exhibits good convergence properties (Fig. 15), though in practice it is still not as powerful as the winding number method. Since the spin stiffness is the free energy change with respect to the twist of the spins components in the $x - y$ plane, it is naturally expected that the estimation of the hydrodynamic relationship should require the model to have U(1) symmetry for this method to be applicable.

Finally, we extracted a high-precision estimate of the velocity in a critical bilayer model and used it in combination with QMC calculated thermodynamic properties to test field-theory predictions based on $1/N$ expansions of the $2 + 1$ dimensional O(3) model. The independently extracted velocity allowed us to precisely determine the deviations of the predicted overall factors in the temperature dependence of the susceptibility $\chi$ and the specific heat $C$. The leading results for the factors obtained in the $1/N$ expansion (which for $\chi$ is independent of $N$ and for $C$ contains a factor $N$), setting $N = 3$, deviate from the correct (QMC computed) values by only about 3%. The Wilson ratio, where the velocity dependence cancels out, agrees perfectly to within the very small (0.3%) error bar (and we cannot establish the level of agreement beyond the level of the error bar).

The remarkably good agreements between the large-$N$ results and the unbiased QMC results are not further improved when the leading corrections to the $N \to \infty$ forms are used. Instead, the agreement becomes significantly worse, with the deviations for $\chi$ and $C$ being about 20% and 10% respectively. In principle there is no reason to expect a better agreement than this, considering that $1/N = 1/3$ is not a very small number, and what is surprising is the very good agreement with the large-$N$ forms without the corrections. In particular, the Wilson ratio, which for the class of SU($N$) systems generalizing the SU(2) system consider here, essentially counts the number of gapless critical modes, $W \propto 1/N$, agrees almost perfectly (with deviations not detectable within the statistical errors of the present study) when the corrections are not included. It is tempting to speculate that the mode-counting property of $W$ is exact, similar to corrections discussed recently for the entanglement entropy [58–60]. An interesting question would then be why the $1/N$ corrections apparently violate the mode-counting property and how this might possibly be avoided.

## APPENDIX: NATURE OF THE SYSTEMATICAL ERRORS IN THE ($n, \beta_{\text{int}}$) GAP ESTIMATOR

We will show analytically the systematic error of the gap estimator (37) and the limits (38) in more details, which was shown in Sec. II B. Let us recall the dynamical correlation

function in the exponential-form projection:

$$
C(\tau, \tau_0) = \frac{1}{Z'} \langle \psi_t | e^{-(\beta - \tau - \tau_0)H} A^\dagger e^{-\tau H} A e^{-\tau_0 H} | \psi_t \rangle
$$
$$
= \frac{1}{Z'} \sum_{\ell, p, q} b_{\ell, p, q}\, e^{-\beta E_p} e^{-\tau(E_\ell - E_p)} e^{-\tau_0(E_q - E_p)},
$$

where $Z' = \langle \psi_t | e^{-\beta H} | \psi_t \rangle$, $|\psi_t\rangle = \sum_q c_q |q\rangle$, $H|q\rangle = E_q |q\rangle$, and $b_{\ell, p, q} = \bar{c}_p c_q \langle p | A^\dagger | \ell \rangle \langle \ell | A | q \rangle$. The Fourier transformed correlation function at frequency $\omega_k = 2\pi k / \beta_{\text{int}}$ ($k \in \mathbf{Z}$) is expressed as

$$
\tilde{C}(\omega_k) = \int_0^{\beta_{\text{int}}} d\tau\, C(\tau, \tau_0) e^{i\tau\omega_k}
$$
$$
= \begin{cases} \frac{1}{Z'} \sum_{\ell, p} g_{\ell, p} \frac{\Delta_{\ell, p} + i\omega_k}{\Delta_{\ell, p}^2 + \omega_k^2} & (\omega_k \neq 0) \\ \frac{1}{Z'} \left( \sum_{E_\ell \neq E_p} \frac{g_{\ell, p}}{\Delta_{\ell, p}} + \sum_{E_\ell = E_p} d_{\ell, p} \beta_{\text{int}} \right) & (\omega_k = 0) \end{cases},
$$
$$
\tag{A1}
$$

where $\beta_{\text{int}} \leqslant \beta - \tau_0$, $g_{\ell, p} = d_{\ell, p}(1 - e^{-\beta_{\text{int}} \Delta_{\ell, p}})$, $d_{\ell, p} = e^{-(\beta - \tau_0)E_p} \sum_q b_{\ell, p, q} e^{-\tau_0 E_q}$, $\Delta_{\ell, p} = E_\ell - E_p$. Note that the imaginary part for $\omega_k \neq 0$ does not vanish as the finite-temperature case because the correlation function $C(\tau, \tau_0)$ is not periodic.

The gap estimator (37) is rewritten as

$$
\frac{\hat{\Delta}_{(n, \beta_{\text{int}})}}{\Delta_1} = \frac{1 + R_n(\beta_{\text{int}})}{1 + F_n(\beta_{\text{int}}) + D_n(\beta_{\text{int}})}
$$
$$
\to 1 + \sum_{\ell > 1} \left\{ \frac{b_\ell}{b_1} \left( \frac{\Delta_1}{\Delta_\ell} \right)^{2n} + O\left[ \left( \frac{\Delta_1}{\Delta_\ell} \right)^{2n+1} \right] \right\}
$$
$$
\times (\beta_{\text{int}} \to \infty), \tag{A2}
$$

where

$$
R_n(\beta_{\text{int}}) = \sum_{\{\ell, p\} \in S} \left( \frac{g_{\ell, p}}{g_{1,0}} \right) \frac{h(n, \beta_{\text{int}}, \ell, p)}{h(n, \beta_{\text{int}}, 1, 0)},
$$
$$
F_n(\beta_{\text{int}}) = \sum_{\{\ell, p\} \in S} \left( \frac{g_{\ell, p}}{g_{1,0}} \right) \frac{h(n, \beta_{\text{int}}, \ell, p)}{h(n, \beta_{\text{int}}, 1, 0)} \left( \frac{\Delta_1}{\Delta_{\ell, p}} \right),
$$
$$
D_n(\beta_{\text{int}}) = \sum_{E_\ell = E_p} x_{n,0,0} \left( \frac{d_{\ell, p}}{g_{1,0}} \right) \beta_{\text{int}} \Delta_1 \frac{1}{h(n, \beta_{\text{int}}, 1, 0)},
$$
$$
h(n, \beta_{\text{int}}, \ell, p) = (-1)^n \omega_1^{2n} \prod_{k=1}^n \frac{1}{\Delta_{\ell, p}^2 + \omega_k^2}
$$

$b_\ell \equiv b_{\ell, 0, 0} = |c_0 \langle \ell | A | 0 \rangle|^2$, and $\Delta_\ell = E_\ell - E_0$. For the summation of $R_n$ and $F_n$, the set $S$ is such that $E_\ell \neq E_p$ except $(\ell, p) = (1, 0)$. Then

$$
\lim_{n \to \infty} \lim_{\beta_{\text{int}}, \tau_0 \to \infty} \hat{\Delta}_{(n, \beta_{\text{int}})} = \Delta_1. \tag{A3}
$$

Let us consider taking the limit $n \to \infty$ before $\beta_{\text{int}}, \tau_0 \to \infty$. Using the product expansion form of the hyperbolic function, $\sinh(\pi z) = \pi z \prod_{k=1}^\infty (1 + z^2/k^2)$, we obtain the limiting form as

$$
\frac{\hat{\Delta}_{(n, \beta_{\text{int}})}}{\Delta_1} = \frac{1 + R_\infty(\beta_{\text{int}})}{1 + G_\infty(\beta_{\text{int}})} + O\left( \frac{1}{n} \right), \tag{A4}
$$

where

$$R_\infty(\beta_{\rm int}) = \sum_{\{\ell,p\}\in S} \left(\frac{d_{\ell,p}}{d_1}\right)\left(\frac{\Delta_{\ell,p}}{\Delta_1}\right)e^{-\frac{\beta_{\rm int}}{2}(\Delta_{\ell,p}-\Delta_1)},$$

$$G_\infty(\beta_{\rm int}) = \sum_{(\ell,p)\neq(1,0)} \left(\frac{d_{\ell,p}}{d_1}\right)e^{-\frac{\beta_{\rm int}}{2}(\Delta_{\ell,p}-\Delta_1)},$$

(A5)

$d_1 \equiv d_{1,0}$, and $G_\infty(\beta_{\rm int}) = \lim_{n\to\infty}(F_n(\beta_{\rm int}) + D_n(\beta_{\rm int}))$. Let us next consider the limit $\beta_{\rm int} \to \infty$. We can rewrite the correction term as

$$G_\infty(\beta_{\rm int}) = \sum_{(\ell,p)\neq(1,0)} \bar{b}_{\ell,p}e^{-(\beta-\tau_0)\Delta_{p,0}-\frac{\beta_{\rm int}}{2}(\Delta_{\ell,p}-\Delta_1)}$$

$$= \sum_{(\ell,p)\neq(1,0)} \bar{b}_{\ell,p}e^{-(\beta-\tau_0-\frac{\beta_{\rm int}}{2})\Delta_{p,0}-\frac{\beta_{\rm int}}{2}\Delta_{\ell,1}}$$

$$\to 0 \quad (\beta_{\rm int}\to\infty),$$

(A6)

where

$$\bar{b}_{\ell,p} = \frac{\sum_q b_{\ell,p,q}e^{-\tau_0 E_q}}{\sum_q b_{1,0,q}e^{-\tau_0 E_q}}.$$

(A7)

Note that $\beta - \tau_0 - \frac{\beta_{\rm int}}{2} \geqslant \frac{\beta_{\rm int}}{2}$ since $\beta \geqslant \tau_0 + \beta_{\rm int}$, $\Delta_{p,0} > 0$ ($p \neq 0$), and $\Delta_{\ell,1} > 0$ ($\ell > 1$). Similarly, $R_\infty(\beta_{\rm int}) \to 0$ ($\beta_{\rm int} \to \infty$). Consequently,

$$\lim_{\beta_{\rm int}\to\infty}\lim_{n\to\infty}\hat{\Delta}_{(n,\beta_{\rm int})} = \Delta_1.$$

(A8)

The convergence of $\beta_{\rm int} \to \infty$ limit is accelerated together with taking $\tau_0 \to \infty$ limit practically. Therefore we have analytically shown that the limits are interchangeable:

$$\lim_{n\to\infty}\lim_{\beta_{\rm int},\tau_0\to\infty}\hat{\Delta}_{(n,\beta_{\rm int})} = \lim_{\beta_{\rm int},\tau_0\to\infty}\lim_{n\to\infty}\hat{\Delta}_{(n,\beta_{\rm int})} = \Delta_1.$$

This appealing property makes the estimation robust in practice and allows for flexibility in how the limit is taken.

As we have shown in the main text, we extrapolate the final gap estimation for $n \to \infty$ first and for $\beta_{\rm int} \to \infty$ later.

It is because data for large $\beta_{\rm int}$ will be noisy, and it is easier to take the limit for $n \to \infty$ first. Nonetheless, we could invert the limits. Then let us consider the finite $n$ correction for making the extrapolation more reliable. Let $Q(n,z) = \prod_{k=1}^n(1 + z^2/k^2)$. We have already used the limit $Q(n,z) \to \sinh(\pi z)/\pi z$ ($n \to \infty$). The finite $n$ correction is expressed as

$$Q(n,z) \sim \frac{\sinh(\pi z)}{\pi z}\exp\left[-\frac{z^2}{n+1}-\frac{z^2}{2(n+1)(n+2)}\right]$$

$$\sim \frac{\sinh(\pi z)}{\pi z}\left[1-\frac{z^2}{n+1}-\frac{z^2}{2(n+1)(n+2)}\right.$$

$$\left.+\frac{z^4}{2(n+1)^2}\right].$$

Here the asymptotic expansion of the Riemann zeta function was used,

$$\sum_{k=1}^n \frac{1}{k^2} = \frac{\pi^2}{6} - \frac{1}{n+1} - \frac{1}{2(n+1)(n+2)} - \cdots.$$

(A9)

As a result, the asymptotic systematic error of the gap estimator is rewritten for $\beta_{\rm int},\tau_0 \gg 1$ as

$$\hat{\Delta}_{(n,\beta_{\rm int})} \sim \Delta_1 + \sum_{\ell>1}\left(\frac{b_\ell}{b_1}\right)\Delta_{\ell,1}e^{-\frac{\beta_{\rm int}}{2}\Delta_{\ell,1}}\left[1 + \frac{\bar{z}_\ell^2}{n+1}\right.$$

$$\left.+\frac{\bar{z}_\ell^2}{2(n+1)}\left(\frac{\bar{z}_\ell^2}{n+1}+\frac{1}{n+2}\right)+O\left(\frac{1}{n^3}\right)\right],$$

(A10)

where $b_\ell > 0$ ($\ell \geqslant 1$), $\Delta_{\ell,1} \equiv \Delta_\ell - \Delta_1 > 0$, and $\bar{z}_\ell^2 = \left(\frac{\beta_{\rm int}}{2\pi}\right)^2(\Delta_\ell^2 - \Delta_1^2) > 0$ ($\ell > 1$). In practice, we can use a fitting function $f(n,\beta_{\rm int}) = \Delta_1 + a\exp(-b\,\beta_{\rm int})(1 + c/n + d/n^2)$ with positive real parameters $a,b,c$, and real parameter $d$ to extrapolate the first gap, as demonstrated in Figs. 5 and 6 in the main text. (Note that the parameter $d$ may be negative for small $\beta_{\rm int}$.)

---

[1] E. Manousakis, Rev. Mod. Phys. **63**, 1 (1991).

[2] K. J. Runge, Phys. Rev. B **45**, 7229 (1992); **45**, 12292 (1992).

[3] U.-J. Wiese and H.-P. Ying, Z. Phys. B **93**, 147 (1994).

[4] A. W. Sandvik, Phys. Rev. B **56**, 11678 (1997).

[5] A. W. Sandvik and H. G. Evertz, Phys. Rev. B **82**, 024407 (2010).

[6] F.-J. Jiang and U.-J. Wiese, Phys. Rev. B **83**, 155120 (2011).

[7] R. R. P. Singh, M. P. Gelfand, and D. A. Huse, Phys. Rev. Lett. **61**, 2484 (1988).

[8] A. J. Millis and H. Monien, Phys. Rev. Lett. **71**, 210 (1993); Phys. Rev. B **50**, 16606 (1994).

[9] A. V. Chubukov and D. K. Morr, Phys. Rev. B **52**, 3521 (1995).

[10] A. W. Sandvik and D. J. Scalapino, Phys. Rev. Lett. **72**, 2777 (1994).

[11] A. W. Sandvik, Phys. Rev. Lett. **83**, 3069 (1999).

[12] R. K. Kaul and R. G. Melko, Phys. Rev. B **78**, 014417 (2008).

[13] F.-J. Jiang, Phys. Rev. B **83**, 024419 (2011).

[14] B. I. Halperin and P. C. Hohenberg, Phys. Rev. **188**, 898 (1969).

[15] S. Chakravarty, B. I. Halperin, and D. R. Nelson, Phys. Rev. B **39**, 2344 (1989).

[16] A. V. Chubukov, S. Sachdev, and J. Ye, Phys. Rev. B **49**, 11919 (1994).

[17] R. K. Kaul, R. G. Melko, and A. W. Sandvik, Annu. Rev. Cond. Matt. Phys. **4**, 179 (2013).

[18] J.-K. Kim and M. Troyer, Phys. Rev. Lett. **80**, 2705 (1998).

[19] P. V. Shevchenko, A. W. Sandvik, and O. P. Sushkov, Phys. Rev. B **61**, 3475 (2000).

[20] M. Matsumoto, C. Yasuda, S. Todo, and H. Takayama, Phys. Rev. B **65**, 014407 (2001).

[21] W. Brenig, Phys. Rev. B **73**, 104450 (2006).

[22] L. Wang, K. S. D. Beach, and A. W. Sandvik, Phys. Rev. B **73**, 014431 (2006).

[23] S. Wenzel and W. Janke, Phys. Rev. B **79**, 014410 (2009).

[24] L. Fritz, R. L. Doretto, S. Wessel, S. Wenzel, S. Burdin, and M. Vojta, Phys. Rev. B **83**, 174416 (2011).

[25] M. Troyer, H. Kontani, and K. Ueda, Phys. Rev. Lett. **76**, 3822 (1996).

[26] A. F. Albuquerque, M. Troyer, and J. Oitmaa, Phys. Rev. B **78**, 132402 (2008).

[27] A. W. Sandvik, V. N. Kotov, and O. P. Sushkov, Phys. Rev. Lett. **106**, 207203 (2011).

[28] H. Suwa and S. Todo, Phys. Rev. Lett. **115**, 080601 (2015).

[29] A. W. Sandvik, Phys. Rev. Lett. **95**, 207203 (2005).

[30] A. W. Sandvik, Phys. Rev. B **59**, R14157 (1999).

[31] S. Liang, B. Doucot, and P. W. Anderson, Phys. Rev. Lett. **61**, 365 (1988).

[32] A. W. Sandvik, J. Phys. A **25**, 3667 (1992).

[33] A. W. Sandvik, AIP Conf. Proc. **1297**, 135 (2010).

[34] M. Jarrell and J. E. Gubernatis, Phys. Rep. **269**, 133 (1996).

[35] A. W. Sandvik, arXiv:1502.06066.

[36] J. Des Cloizeaux and J. J. Pearson, Phys. Rev. **128**, 2131 (1962).

[37] M. Takahashi, *Thermodynamics of One-Dimensional Solvable Models* (Cambridge University Press, Cambridge, UK, 1999).

[38] J.-S. Caux (private communication).

[39] E. Farhi, D. Gosset, I. Hen, A. W. Sandvik, P. Shor, A. P. Young, and F. Zamponi, Phys. Rev. A **86**, 052334 (2012).

[40] C. De Grandi, A. Polkovnikov, and A. W. Sandvik, Phys. Rev. B **84**, 224303 (2011).

[41] B. B. Beard and U.-J. Wiese, Phys. Rev. Lett. **77**, 5130 (1996).

[42] N. V. Prokofév, B. V. Svistunov, and I. S. Tupitsyn, Zh. Eksp. Teor. Fiz. **114**, 570 (1998) [Sov. Phys. JETP **87**, 310 (1998)].

[43] A. W. Sandvik, R. R. P. Singh, and D. K. Campbell, Phys. Rev. B **56**, 14510 (1997).

[44] V. Neagoe, IEEE Signal Process. Lett. **3**, 119 (1996).

[45] B. A. Berg, *Markov Chain Monte Carlo Simulations and Their Statistical Analysis* (World Scientific, Singapore, 2004).

[46] E. L. Pollock and D. M. Ceperley, Phys. Rev. B **36**, 8343 (1987).

[47] A. W. Sandvik and J. Kurkijärvi, Phys. Rev. B **43**, 5950 (1991).

[48] M. Suzuki, S. Miyashita, and A. Kuroda, Prog. Theor. Phys. **58**, 1377 (1977).

[49] J. E. Hirsch, R. L. Sugar, D. J. Scalapino, and R. Blankenbecler, Phys. Rev. B **26**, 5033 (1982).

[50] S. Yasuda, H. Suwa, and S. Todo, Phys. Rev. B **92**, 104411 (2015).

[51] C. J. Hamer, G. R. W. Quispel, and M. T. Batchelor, J. Phys. A **20**, 5677 (1987).

[52] S. Eggert, I. Affleck, and M. Takahashi, Phys. Rev. Lett. **73**, 332 (1994).

[53] T. M. R. Byrnes, R. J. Bursill, H.-P. Eckle, C. J. Hamer, and A. W. Sandvik, Phys. Rev. B **66**, 195313 (2002).

[54] C. M. Canali, S. M. Girvin, and M. Wallin, Phys. Rev. B **45**, 10131 (1992).

[55] H. Neuberger and T. Ziman, Phys. Rev. B **39**, 2608 (1989); D. S. Fisher, *ibid.* **39**, 11783 (1989); P. Hasenfratz and F. Niedermayer, Z. Phys. B **92**, 91 (1993).

[56] O. F. Syljuasen and A. W. Sandvik, Phys. Rev. E **66**, 046701 (2002).

[57] B. Bauer *et al.*, J. Stat Mech. (2011) P05001.

[58] A. B. Kallin, E. M. Stoudenmire, P. Fendley, R. R. P. Singh, and R. G. Melko, J. Stat. Mech. (2014) P06009.

[59] M. A. Metlitski and T. Grover, arXiv:1112.5166.

[60] B. Kulchytskyy, C. M. Herdman, S. Inglis, and R. G. Melko, Phys. Rev. B **92**, 115146 (2015).